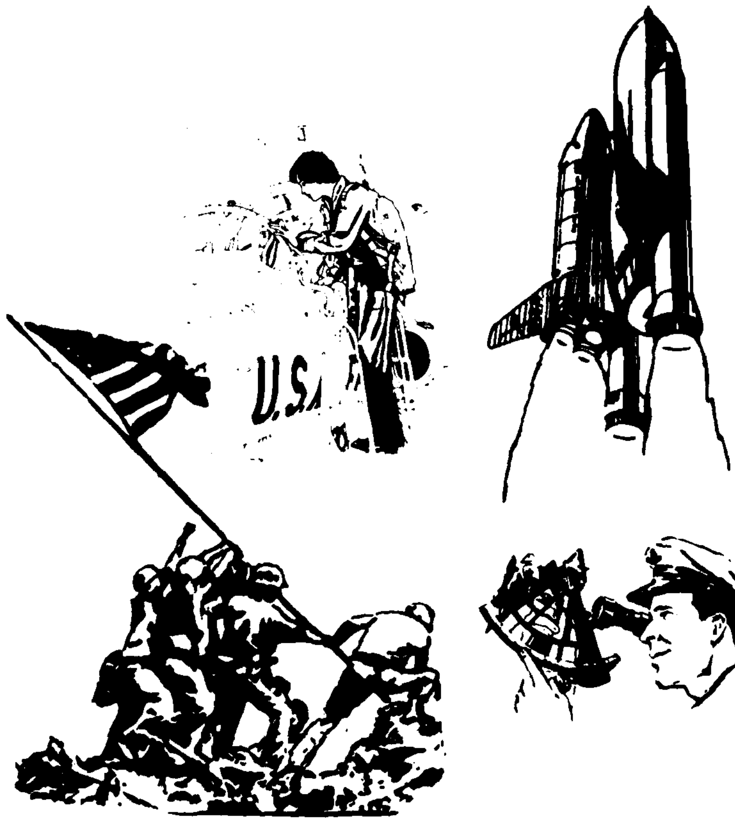
USAFA-TR-86-1

PROCEEDINGS PSYCHOLOGY IN THE DEPARTMENT OF DEFENSE

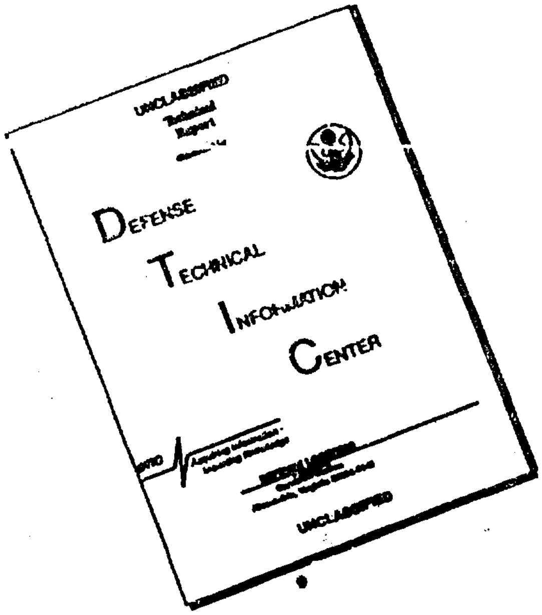


TENTH SYMPOSIUM

16-18 APRIL 1986

SPONSOR:
USAF ACADEMY DEPARTMENT OF BEHAVIORAL SCIENCES AND LEADERSHIP

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

ADA 166570

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release and Distribution Unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) USAFA TR 86-1	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Dept of Behavioral Sciences 6b. OFFICE SYMBOL (If applicable) DFBL	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State and ZIP Code) U.S. Air Force Academy Colorado Springs, CO 80840-5941	7b. ADDRESS (City, State and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Office of Scientific Research 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State and ZIP Code)	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT NO.
11. TITLE (Include Security Classification) Proceedings, Psychology in the Department of Defense, Tenth Annual Symposium (UNCLASSIFIED)	
12. PERSONAL AUTHOR(S) GEORGE E. LEE, Maj, USAF	
13a. TYPE OF REPORT Proceedings 13b. TIME COVERED FROM ______ TO ______	14. DATE OF REPORT (Yr., Mo., Day) April 1986 15. PAGE COUNT 691
16. SUPPLEMENTARY NOTATION	
17. COSATI CODES FIELD / GROUP / SUB. GR.	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Psychology, Symposium, Training, Leadership, Radiation Effects, Counseling, Cognitive Functioning, Operator Performance, Assessment, Stress, Job Performance, Turnover, Workload, Familiy Issues, Biofeedback, Retention.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

These printed proceedings include papers and presentation that deal with a wide range of research in psychology with emphasis on military issues.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL GEORGE E. LEE, Maj, USAF	22b. TELEPHONE NUMBER (Include Area Code) (303) 472-4174 22c. OFFICE SYMBOL DFBL

DD FORM 1473, 83 APR

EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

FOREWORD

The opinions expressed in this volume are those of the individual authors and do not necessarily reflect official policy either of the Army, Navy, Air Force, Coast Guard or of any government organization in which the author(s) may be serving.

Trade names of materials or products of commercial or non-government organizations are cited only where essential to precision in describing research procedures or evaluation of results. Their use does not constitute official indorsement or approval of the use of such commercial hardware or software.

These proceedings are approved for public release, distribution unlmited, in accordance with paragraph IVa, AFR 80-45.

In publishing the Proceedings of the Tenth symposium, the editor sought to facilitate an effective and timely dissemination of the technical information presented. The papers contained in this document were printed directly from unedited reproducible copy submitted by the authors who are solely responsible for their contents.

Acknowledgements

Cover: Graphics Division of Audiovisual Services
Dean of Faculty, USAF Academy

Editorial Assistance:

Janet Gomes
Erika Hirt
Karla Sanchez
Kathy Rockefeller

PROCEEDINGS OF THE TENTH SYMPOSIUM
PSYCHOLOGY IN THE DEPARTMENT OF DEFENSE
16-18 April 1986

Department of Behavioral Sciences and Leadership
United States Air Force Academy
Colorado Springs, Colorado

Symposium Chairman
Lieutenant Colonel William L. Derrick

EDITOR

Major George E. Lee

SYMPOSIUM COMMITTEES

Professor and Acting Head, Department of
Behavioral Sciences and Leadership

Lieutenant Colonel Richard L. Hughes

PROGRAM COMMITTEE

Lieutenant Colonel Robert C. Ginnett, Chair
Major Kevin W. O'Callaghan
Major George E. Lee
Captain John Micalizzi
Captain Lois Short

INVITATIONS AND REGISTRATION

Major Randolph J. Chalkley, Chair
Major William P. Marshak
Major Gregory D. Smith

SOCIAL, BILLETING & TRANSPORTATION COMMITTEE

Major Francis E. McIntire, Jr., Chair
Major Alfred J. Guardino
Captain Mark W. Patton
Captain Melissa R. Kallett

FACILITIES AND VISUAL AIDS COMMITTEE

Captain Gordon J. Curphy, Chair
Staff Sergeant David A. White
Staff Sergeant Robert L. Stewart

PROTOCOL COMMITTEE

Captain Michael S. Williams, Chair
Captain Jeffrey S. Austin
Captain Ralph S. Hartman

PUBLICITY COMMITTEE

Captain Joy L. Prosise, Chair
Captain Ann C. Musselman

This research report is presented as a competent treatment of the subject, worthy of publication. The United States Air Force Academy vouches for the quality of the research, without necessarily endorsing the opinions and conclusions of the authors.

This report has been cleared for open publication and/or public release by the appropriate Office of Information in accordance with AFR 190-17 and DODD 5230.0. There is no objection to unlimited distribution of this report to the public at large, or by DDC to the National Technical Information Service.

This research report has been reviewed and is approved for publication.

ROBERT L. JAMES, Major, USAF
Director of Research and
Computer Based Education

Additional copies of this document are available through the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Road, Springfield, VA 22151.

PROCEEDINGS OF THE TENTH SYMPOSIUM ON PSYCHOLOGY IN THE DEPARTMENT OF DEFENSE

PANEL SESSION: AN AUTOMATED CLASSROOM FOR ARMOR TRAINING

PANEL SESSION: RECENT DEVELOPMENTS IN TEAM TRAINING AND PERFORMANCE MEASUREMENT

Discussant: Bert King, Office of Naval Research

PAPER SESSION: LEADERSHIP

PANEL SESSION: STRESS AND PERFORMANCE

PAPER SESSION: VISUAL PROCESSING II

PAPER SESSION: CAREERS III: STUDIES OF ATTRITION AND RETENTION

PAPER SESSION : WORKLOAD

PANEL SESSION: BEHAVIORAL ISSUES ASSOCIATED WITH ISOLATION AND CONFINEMENT

PAPER SESSION: VALUES, ATTITUDES, AND OPINIONS

PANEL SESSION: TEACHING AIR FORCE LEADERSHIP IN THE CLASSROOM

Panelists: Richard Lester, Paul R. Murphy, James Mullett, David L. Herres (Maxwell AFB, AL)
Bernard Krockover (Lackland AFB, TX)
Lynn Scott (U.S. Air Force Academy)

PAPER SESSION: JOB PERFORMANCE: PREDICTION, ASSESSMENT, AND IMPROVEMENT

Performance and Preference under
Chromatic Ambient Illumination

S.M. Luria, David A. Kobus,[1] David F. Neri
Naval Submarine Medical Research Laboratory

Abstract

Five different ambient illuminations are found on U.S. submarines-- bright and dim white, blue, red, and no ambient light. A series of studies comparing performance and preference under the various lights is described. Performance does not always correlate with preference, and preferences change under different conditions. Dim white light is recommended for night-time ambient illumination.

During World War II red ambient light began to be used to illuminate submarine compartments at night, because it facilitated dark adaptation. But it has always been unpopular. Many complained that it led to eye fatigue and made it impossible to read color-coded charts.

The fact that submariners no longer stand deck watches every night, the increased use of color-coded panels, and the imminent use of color-coded cathode-ray tube displays have now led submariners to question whether red light is still needed.

A few years ago one sonar crew substituted blue ambient light for the red. They contended that the blue light resulted in better visibility of their CRT displays, less fatigue, and better performance. No testing had been done, however, and there was, therefore, no real evidence of improved vision or performance. Since good night vision is still required of the periscope operators, and since they occasionally go into the sonar room to look at the displays, it could be argued that blue was a poor choice. White light, equated in brightness to the red, would interfere less with dark adaptation while improving vision. We therefore began a comparative evaluation of different ambient illuminants.

Since improved contrast of the displays had been reported, we first measured contrast sensitivity for gratings displayed on a CRT under red, white, and blue ambient illumination matched photopically for brightness at CRT intensity levels of 1 and 0.1 fL (Neri & Kinney, 1982). There were no differences in sensitivity under the three lights (Fig.1),

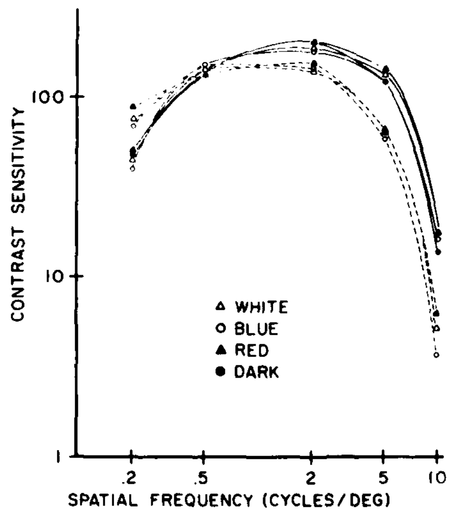


Fig. 1. Contrast sensitivity under four colors of ambient illumination and screen brightnesses of 1 (solid lines) and 0.1 (broken lines) fL.

[1] Now at Naval Health Research Center.

and we concluded that the preference for blue light was not due to any enhancement of contrast or contrast sensitivity.

We next investigated the ability to detect actual targets--presented in the submarine school sonar simulator--under red, white, and blue light. Experienced crews regularly come for refresher training lasting several days. We tested three crews, each receiving the lights in a different order (Kinney, Luria, & Ryan, 1982). Detection ranges were about the same under the different colors.

We next carried out an extensive comparison of the performance and preferences of 20 sonar crews in the sonar simulator under blue and white light (Luria & Kobus, 1983). The lights were equated for brightness (Kinney, 1983). Nearly all these crews trained for four days, allowing us to present the two lights in ABBA order. At the end of each day, we obtained the detection times and ranges for the sonar problems as well as preference ratings of the light. Of the 200 men tested, two-thirds preferred the white light to the blue. Moreover, both detection and classification times were faster under the white, particularly in the difficult baffle area. Five of these crews subsequently evaluated the white light at sea. Four of them preferred the white light. Most interesting, two of the crews which had preferred the blue in the simulator now rated the white light superior, saying it was preferable for the long periods on patrol.

In another study, we compared all five ambient lighting conditions found on submarines (Kobus & Neri, 1984). Using the "waterfall" display found on every submarine (Fig. 2), we measured the times required to detect targets by 20 subjects under each of the five lights in counterbalanced order. Mean detection time was shortest for the dim white and no light conditions (Fig.3). These subjects, however, preferred the red and blue to the dim white and no light, despite the fact that performance was worse under the colored lights.

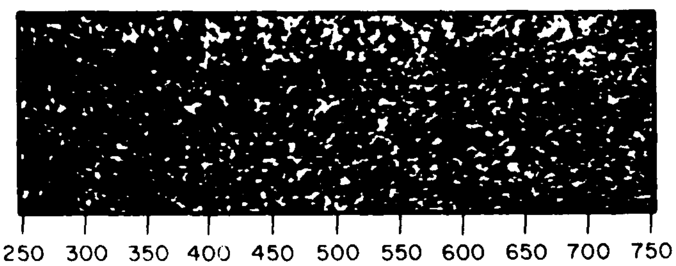


Fig. 2. Waterfall display with a target beginning to appear at 400 Hz.

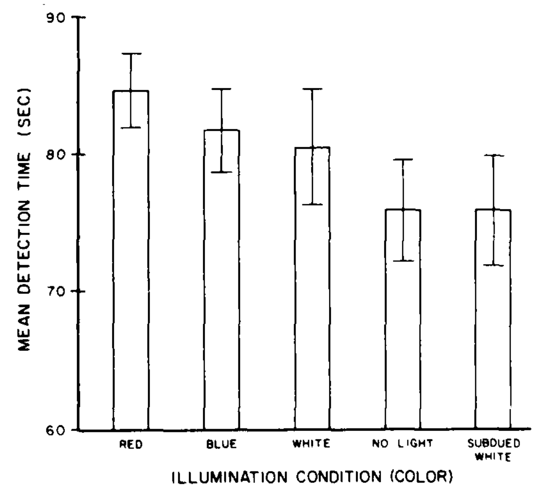


Fig. 3. Detection time under each ambient color. Error bars are ± 1 S.E.M.

We have also attempted to demonstrate the amount of visual fatigue under the various colors of light through measurements of the eye movements. Four emmetropes and four hyperopes searched for targets in simulated sonar displays for an hour under red, blue, or white light (Kinney at al,1983). There were no systematic differences in

the number or duration of their blinks, or in the number of saccades. There were, however, increases in saccade duration and magnitude under red light (Fig. 4) and decreases in saccade velocity for the hyperopes.

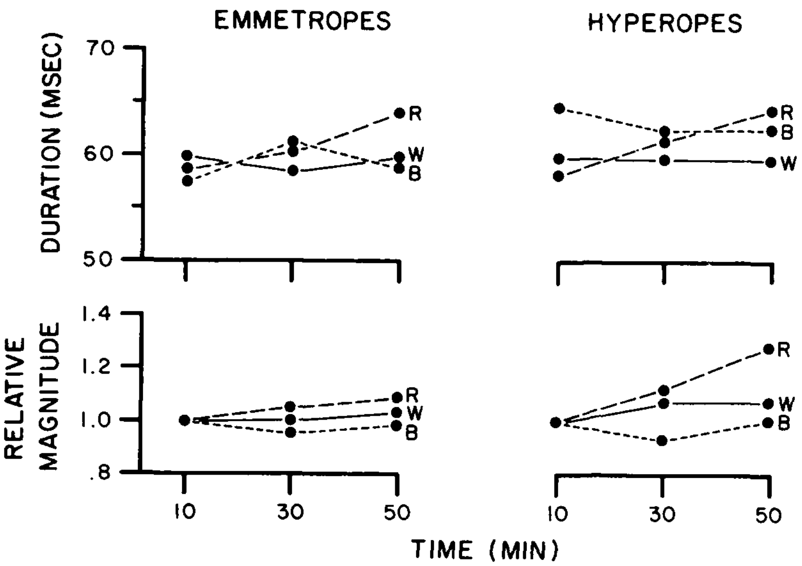


Fig. 4. (Top) Mean saccade duration over time under red (- - - -), white (____), and blue (----) ambient light.

The analysis of saccade velocity is complicated, since larger saccades are faster than smaller ones; and, as noted, saccade magnitude changed. The relationship is a power function which were fitted to the data for the first and last sessions. Figure 5 shows the exponents of the functions and, hence, the relative saccade velocities. This increased slightly for the emmetropes under all colors, but for the hyperopes, there was an appreciable decrease under red as well as a small decrease under white.

Since the adoption of chromatic CRT displays appears imminent, we investigated the effects of chromatic ambient illumination on the ability to discriminate colors on such displays (Neri et al, in press). Circular targets of different colors were presented on backgrounds of different colors. The CRT screen was divided into quadrants, and the subject had to identify which contained the target. The chromatic ambient illuminations, which were matched to the brightness of the red on submarines, had no effect on reaction time to the CRT colors (Fig. 6).

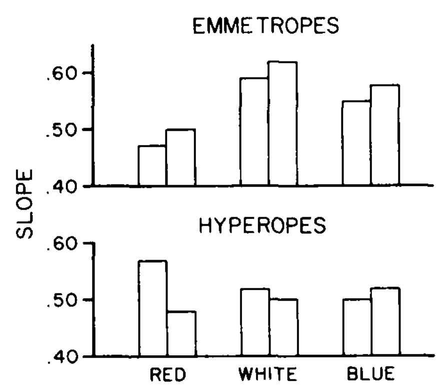


Fig. 5. Slope of the function relating saccade velocity to saccade magnitude in the first and last sessions for emmetropes (top) and hyperopes (bottom). A higher number indicates a higher saccade velocity.

Finally, we measured the ability of observers, who were adapted to either red or white light of equal brightness, to see simulated ship silhouettes and shadows immediately after the ambient light had been extinguished. As soon as it was, the subject looked at a screen illuminated to the level of the night-time sky through a shutter

which automatically closed in 10 seconds. The threshold contrast of these targets was the same after adaptation to low levels of either red or white. One experiment was carried out in the submarine school navigation trainer. Subjects looked through a periscope at an approaching ship for 10-second periods until they could detect it. Detection ranges were again the same after red and white adaptation.

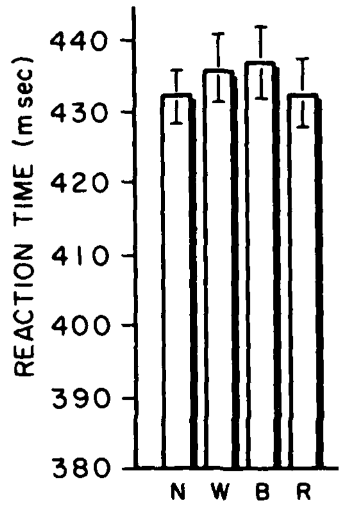


Fig. 6. Mean reaction time under each ambient color. Error bars are 1 S.E.M. N= no light; W=white; B= blue; R= red.

These studies indicate that dim white is preferable to red or blue light for night-time ambient illumination on submarines. The advantage which red light gives for subsequent dark adaptation is more than offset by its disadvantages. And, indeed, a careful consideration of the studies on which the selection of red light was based shows that its advantage has been considerably overstated (Luria and Kobus, 1985b). Although there is no question that dark adaptation is faster after exposure to red than to any other color, this temporal advantage decreases with the intensity of the adapting light. The results shown in Fig. 7 are completely typical. At an adapting intensity of 300-400 mL, red may give a 10-15 minute advantage, but at an adapting intensity of 3 mL, that advantage is reduced to a couple minutes. And ambient intensities on submarines at night are even lower. Hulburt (1951) summarized such experiments in Fig. 8. The one minute advantage conferred by dim red compared to dim white light is hardly of operational significance. The use of red light should be ended.

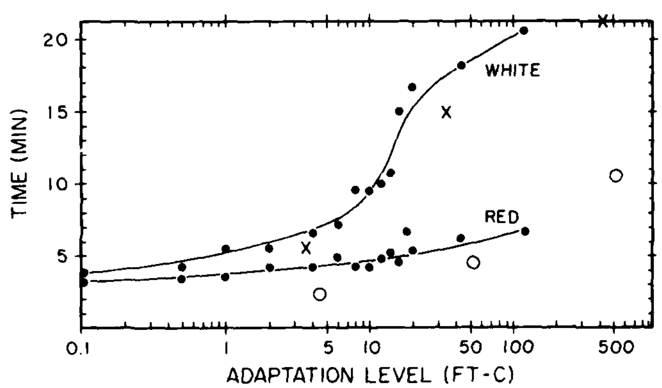


Fig. 8. Time required to dark-adapt after exposure to red or white light of different intensities. (From Hulburt, 1951).

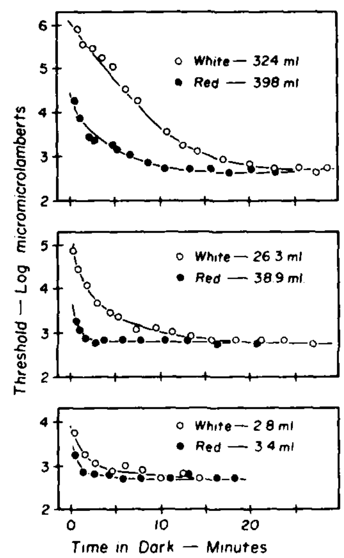


Fig. 7. Dark adaptation after exposure to white or red at three brightness levels. (From Hecht and Hsia, 1945).

Note

The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, the Department of Defense, or the U. S. Government.

References

Hecht, S., & Hsia, Y. (1945). Dark adaptation following light adaptation to red and white lights. Journal of the Optical Society of America, 35, 261-167.

Hulburt, E. (1951). Time of dark adaptation after stimulation by various brightnesses and colors. Journal of the Optical Society of America, 41, 402-404.

Kinney, J. A. S. (1983) The quantity of light produced by red and blue filters over light fixtures in sonar control rooms (Report No. 995). Groton, CT: Naval Submarine Medical Research Laboratory.

Kinney, J. A. S., Luria, S. M., & Ryan, A. P. (1982). Subjective preferences and detection ranges in sonar control rooms under red and blue light (Report No. 991). Groton, CT: Naval Submarine Medical Research Lab.

Kinney, J. A. S., Neri, D. F., Mercado, D. T., & Ryan, A. P. (1983). Visual fatigue in sonar control rooms lighted by red, white or blue illumination (Report No. 1000). Groton, CT: Naval Submarine Medical Research Lab.

Kobus, D. A., & Neri, D. F. (1984). Performance and preference on a sonar detection task under various colors of ambient illumination (Report No. 1023). Groton, CT: Naval Submarine Medical Research Laboratory.

Luria, S. M., & Kinney, J. A. S. (1983). Preferences for blue and white light in sonar compartments (Report No. 1013). Groton, CT: Naval Submarine Medical Research Laboratory.

Luria, S. M., & Kobus, D. A. (1985). Immediate visibility after red and white adaptation (Report No. 1045). Groton, CT: Naval Submarine Medical Research Laboratory.

Luria, S. M., & Kobus, D. A. (1985b). The relative effectiveness of red and white light for subsequent dark-adaptation (Report No. 1036). Groton, CT: Naval Submarine Medical Research Laboratory.

Neri, D. F.,& Kinney, J. A. S. (1982). Contrast sensitivity measured in low levels of red, white, and blue ambient illumination (Report No. 989). Groton, CT: Naval Submarine Medical Research Laboratory.

Neri, D. F., Luria, S. M., & Kobus, D. A. (in press). The detection of various color conditions with different chromatic ambient illuminations. Aviation, Space and Environmental Medicine.

The Influence of Flight Conditions on the Perception of Color CRT Displays

Mary M. Lloyd
Douglas J. Funke
Arvin/Calspan Corporation, Buffalo, New York

Abstract

Based on a review of relevant research, general guidelines were developed for the use of color in military aircraft CRT displays. The guidelines reflect human perceptual capabilities and perceptual phenomena related to environmental conditions (e.g., Purkinje shift, red-out). These guidelines are presented along with supporting rationale.

Introduction

The perception of color can be influenced by various flight conditions, such as increased gs or changes in luminance levels. This is particularly true for military aircraft since conditions of use are more severe than those of civilian aircraft. These perceptual variations may serve to degrade pilot performance and, therefore, must be considered when determining how to apply color in the military cockpit.

LIGHTING CONDITIONS

Night Vision Goggles: Night vision goggles are used to increase the visibility of objects in the outside world by intensifying the light reflected from an object. Since the visibility of objects in the outside world may be decreased by instrument lighting and by reflected light in the the cockpit, dim blue-green lighting may be used for instrument lighting and a red filter may be fitted to the night vision goggles (Wanstall, 1985).

The reason for using blue-green lighting in the cockpit is related to a phenomenon called the "Purkinje shift". When the luminance level decreases from daylight conditions (i.e., photopic vision where both rods and cones function) to night conditions (scotopic vision where only the rods function), the eyes' sensitivity to light waveforms changes. At high levels of illumination, the eye is most sensitive to the color green (550 nm). However, at lower levels of illumination, the eye is most sensitive to the color blue-green (500 nm) (McCormick and Sanders, 1982). It is this change in sensitivity to which the Purkinje shift refers. Hence, the use of blue-green instrument lighting increases the visibility of the displays during night maneuvers. While blue-green lighting enhances visual acuity in a low-luminance environment, red lighting aides in the detection of objects outside the cockpit as it is associated with a lower luminance threshold. An important issue to be resolved when red or blue-green lighting is used is to ensure that surface colors on the instrument displays remain discriminable under these light conditions.

Laser Filters: Laser filters are used to protect the aircrews' eyes from damage inflicted by laser light. Since laser filters selectively decrease the intensity of light, the luminance level of CRT displays must be increased when the filters are used.

Laser filters can also differentially affect the perception of color. For example, when Holst (1976) examined the affect of laser filters on the detectability of low contrast targets, he found that more contrast was needed to detect blue targets (either

natural or rendered blue by laser protective material). However, with high contrast, no degradation of detection performance was found.

Ambient Light: Ambient light conditions can influence color perception. The approximate luminance levels at which surface colors become difficult to discriminate and color confusions may occur are below 3 cd/m^2 and above 3000 cd/m^2 (Bishop and Crook, 1961; cited in Krebs, Wolf, and Sandvig, 1978).

The time to detect light signals will also vary as a function of ambient illumination. For example, under low ambient conditions, signals in the blue to green wavelength region will be detected faster than wavlengths in the red region (Pollack, 1968, cited in Krebs et al., 1978). Recall that the fast detection of blue to green light under low luminance conditions is consistent with the Purkinje Shift. At low levels of luminance the eye is most sensitive to the color blue-green and least sensitive to the color red. Under high ambient illumination signals in the red and the blue wavelength regions will be detected more quickly than those in the yellow or yellow-orange end of the spectrum (Tyte, Wharf and Ellis, 1975, cited in Krebs et al., 1978).

CRT displays with adjustable luminance controls can be used to alleviate some of the problems that ambient light conditions create. Silverstein and Merrifield (1984, p. 45) suggest that "For color display purposes, good color perception and color discrimination can be achieved within the range of 1-1000 fL" (3.43 - 3426 cd/m^2).

Surface Reflectance: Incident ambient light increases the brightness of the display surround which serves to reduce the brightness contrast and color contrast of the display characters. As the visibility and perceived saturation decrease, perception of color and form becomes increasingly difficult. For example, under high ambient light conditions, the color blue looses more visual acuity than other colors (Krebs, et al). Therefore, Krebs et al. (1978) advise that when the color blue is used for alphanumerics or for coding symbols, the display should be carefully tested to ensure that it is legible.

Contrast Enhancement: Variables that influence the degree to which the legibility of a color CRT is degraded by ambient light, include:

1. Color contrast: The basic colors recommended by Krebs et al. (1978) for use on CRT displays are red, yellow and green, since they are highly discriminable from each other. The use of white tends to be confused with the color yellow and the color blue results in poorer legibility than red, yellow or green.

2. Luminance contrast: For optimal contrast symbol-to-background luminance ratio of approximately 10:1 should be used (Haeusing, 1976, cited in Krebs et al., 1978). The luminance ratio recommended for optimal performance may change based on other factors, such as symbol size and number of colors used in the display (Krebs et al., 1978).

   Contrast enhancement filters can be used to prevent bright ambient light from being reflected back from the display surface. It should also be cautioned that loss of contrast enhancement may result through the discoloration of filters that is caused by the ultra-violet rays in sunlight (Yarger and Waruszewski, 1980).

3. Light vs dark CRT background (highly reflective surface vs. the Black Hole effect): There is some controversy surrounding the use of a light or dark display background. For example, Krebs et al. (1978) recommend the use of

a dark background since color symbols are more visible on dark backgrounds. Other researchers, however, have recommended that light display backgrounds result in better performance. They claim that when a dark background and dark surround are used, discrepant depth perception cues from the various colors may cause luminous symbols to appear to float (Farrell and Booth, 1975, cited in Silverstein and Merrifield, 1984). According to Farrell and Booth (1975, cited in Silverstein and Merrifield, 1984), use of a lighter surround increases sensitivity to color since color symbols are perceived as more saturated when presented on a light background as compared to a dark one. Therefore, a light background generally serves to facilitate color discrimination and reduces the likelihood of possible color confusions. It is cautioned, however, that the background intensity should not exceed 1.0 fL (3.43 cd/m^2) (Silverstein and Merrifield, 1984).

4. Reflectance of color off the windshield and the glass canopy: Glare occurs when lights within the cockpit are reflected off the windshield and the canopy. Lights and lighted surfaces should be positioned such that reflections will not reach the eye and a minimum luminance level should be used while still maintaining acceptable display legibility.

<u>Flash Luminance/Blindness:</u> Flash blindness may be caused by lightning, an incendiary explosion or a nuclear flash. A Thermal Flash Blindness Protective Device (TFPD) can be used to screen out harmful rays. In its transparent state, the pane allows for a light transmission level of about 20% (Ivey, 1984). Hence, the luminance level of CRT displays must be increased to ensure display legibility when the protective goggles are being used.

Ivey (1984) has investigated the influence of goggles on CRT symbol legibility during both static and low vibration conditions. He found that although there was some degradation of legibility when the goggles were worn in the vibrations condition, error rates and response times remained relatively low. It was suggested that a symbol size greater than 0.12 inches be used when low level vibration conditions are expected to occur.

Flash blindness may differentially influence color perception. Smith and Goddard (1967) cite numerous studies which demonstrated that after exposure to high intensity light (e.g., a close lightning strike) performance for "deep red" was worse than for any other color. Bernberg (1962, cited in Smith and Goddard, 1967) found that subjects took longer to read dials or displays that were illuminated in red light as compared to when they were illuminated in white light.

Richey, Allen and Bower (1980) recommend that when a nuclear flash occurs the luminance level should be increased to a minimum of 7 mL and to a maxium of 100 mL (22.28 to 318.3 cd/m^2). Increasing the luminance of the displays, as suggested, would also increase the time it would take to recover vison for objects outside the cockpit, but it would shorten the recovery time required to read the displays.

LOCATION OF FIELD OF VIEW

<u>Foveal and Peripheral Vision Limitations:</u> The eye is differentially sensitive to colors that occur in the periphery. Sensitivity to the colors yellow and blue extend farther into the periphery than for red and green. However, there seems to be some discrepancy in the literature as to whether sensitivity to blue or to yellow extends further into the periphery.

Response time to color signals also varies within the total field of view (FOV). For example, response time is the shortest for white and it has the widest field of view over the entire field. The color red has the longest response time and the narrowest sensitivity limit. The sensitivity limits and response times for the colors blue, green and yellow fall between white and red parameters (Haines, Dawson, Galvan and Reid, 1975, cited in Krebs, et al, 1978).

Symbol Size for Foveal and Peripheral Displays: Color perception may be impaired when color symbols subtend less than 15 minutes visual arc since the symbols may appear to be less saturated and may shift in apparent hue relative to larger symbols (Burnham, Hanes and Bartleson, 1963; Farell and Booth, 1975, both cited in Silverstein and Merrifield, 1984).

VIBRATION

Visual acuity can be significantly degraded by vibration, especially under conditions of vertical movement. Therefore, vibration presents a more serious problem in the design of helicopter displays as compared to fixed-wing aircraft displays.

In a recent study, the influence of low level vibration on the readability of a color CRT display was examined using the motion simulator (Ivey, 1984). Measures of chrominance (i.e., the qualities of color associated with hue and saturation) and luminance were compared among conditions of low vibration. It was found that of the seven colors (i.e., green, red, yellow, orange, cyan, brown, and white) tested, chrominance and luminance measures for the CRT remained constant during vibration. On the basis of the reaction time data, it was recommended that a symbol size of at least 0.12 inches (0.3 cm) be used when conditions of low level vibration are anticipated. However, under conditions of high level vibration significantly larger symbol size may be required (e.g., from 22.5 to 67.5 arc min).

ACCELERATION CONSIDERATIONS ("G" FORCES)

Gray-out: With the building of high gs, blood pressure in the head decreases. According to Preble (1985), at about +4 gs this lowering of blood pressure can result in tunnel vision or the loss of peripheral vision. The pilot may also experience gray-out, a state in which he or she is still conscious but has suffered a severe loss of visual acuity and the ability to perceive color.

Protective measures that can be taken include the use of "g-suits" and straining manuevers (i.e., the tensing of muscles to keep blood from flowing downward, away from the head). During the straining manuevers, the pilot may be unable to read displays. Reclining seats are also used to increase g tolerance. However, the use of reclining seats places constraints on where the CRT displays are located as the pilot's unobstructed view is extremely narrow when the seat is angled back (Roe, 1981).

Red-out: According to McFarland (1946), the body's ability to withstand negative acceleration is less than its tolerance to positive g force. At levels of -2 to -3 gs, the pilot may experience head pain and "red-out". Red-out, or the reddening of vision, results when the vessels in the eye become abnormally full of blood.

Hypoxia: Hypoxia is the state of oxygen deprivation. Hypoxia can impair dark adaptation and it affects visual sensitivity the most in dim illumination. McFarland (1946) likens the effect of oxygen deprivation on vision to that of placing a piece of smoked glass between the eye and the light source. An increase in light intensity will serve to lessen this loss of sensitivity (McFarland, 1946).

REFERENCES

Beilman, J.L., Funke, D.J., Gawron, V.J., Hammill, H.B., Lloyd, M.M. and Mitchell, J.F. (1985, September). Guidelines for Using Airborne Color Displays. Buffalo, New York: Arvin/Calspan Corporation.

Bernberg, R.E. A Comparison of Response Latencies for Visual Recovery from Flash in Red and White Ambient Illumination. Journal of Engineering Psychology, 1, 1-11.

Bishop, H.P., and Crook, M.N. (1961, March). Absolute Identification of Color for Targets Presented Against White and Colored Backgrounds. (Report No. WADD-TR-60-611). Wright-Patterson AFB, Ohio: Author.

Burnham, R.W., Hanes, R.M., and Bartleson, C.J. (1963). Color: A Guide to Basic Facts and Concepts. New York: John Wiley and Sons, Inc.

Farrell, R.J., and Booth, J.M. (1975). Design Handbook for Imagery Interpretation Equipment. Seattle, Washington: Boeing Aerospace Company.

Haeusing, M. (1976). Color Coding of Information on Electronic Displays. Proceedings of the Sixth Congress of the International Ergonomics Association, pp. 210-217.

Haines, R.M., Dawson, L.M., Galvan, T., and Reid, L.M. (1975, March). Response Time to Colored Stimuli in the Full Visual Field. (Technical Report NASA-TN-D-7927). Moffett Field, CA: NASA Ames Research Center.

Holst, G.C. (1976, April). Reduction of Target Detectability by Laser Protective Materials. (Technical Report FA-TR-76021). Philadephia, PA: Frankford Arsenal.

Ivey, L. (1984, December). Effect of Vibration on the Readability of Color CRT Displays. In Third Aerospace Behavioral Engineering Technology Conference Proceedings. "Automation Workload Technology: Friend or Foe?" (pp. 53-59). Warrendale, PA: Society of Automobile Engineers.

Krebs, M.J., Wolf, J.D., and Sandvig, J.H. (1978, October). Color Display Design Guide. (Report ONR-CR213-136-2F). Minneapolis, Minnesota: Honeywell Systems and Research Center.

McCormick, E.J., and Sanders, M.S. (1982). Human Factors in Engineering and Design (Fifth ed.). New York: McGraw-Hill Book Company, Inc.

McFarland, R.A. (1946). Human Factors in Air Transport Design. New York: McGraw-Hill Book Company, Inc.

Pollack, J.D. (1968). Reaction Time to Different Wavelengths at Various Luminances. Perception and Psychophysics, 3, 17-24.

Preble, C. (1985, March 11). USAF Developing Electronic Valve to Improve Resistance to High G-Forces. Aviation Week and Space Technology, 122 (10), 50-52.

Richey, E.O., Allen, R.G., and Bower, J.L. (1980, December). Evolution of PLZT Goggles. (Technical Report SAM-TR-80-17). Brooks Air Force Base, Texas: USAF School of Aerospace Medicine.

Roe, G. (1981). Human Engineering in Aircraft and System Design. In Advanced Aircrew Display Symposium (5th), 4-38.

Silverstein, L.D., and Merrifield, R.M. (1984). The Development and Evaluation of Color Display Systems for Airborne Applications. Seattle, Washington: Boeing Commercial Airplane Company.

Smith, H.A., and Goddard, C. (1967, May). Effects of Cockpit Lighting Color on Dark Adaptation. (Technical Report AFDL-TR-67-56). Wright-Canoga Park, California: The Bunker-Ramo Corporation.

Tyte, R., Wharf, J., and Ellis, B. (1975). Visual Response Times in High Ambient Illumination. Society of Information Displays Digest, 98-99.

Wanstall, B. (1985). Among the Weeds at Night: Transition to Night NOE Flying. Interavia, 40 (7), 793-795.

Yarger, Keith E., Waruszewski, Harry L. (1980, May). Effects of Sunlight on Display Filters. In Proceedings of the IEEE 1980 National Aerospace and Electronics Conference, 3, 1216-1218, Dayton, Ohio: IEEE.

Guidelines For Color Coding In Cockpits

Douglas J. Funke
Mary M. Lloyd
Arvin/Calspan Corporation, Buffalo, New York

Abstract

General guidelines for color coding in cockpits were developed through a review of relevant literature. A summary of these guidelines includes when to use color, the number of colors to use, assigning meaning to colors and color pitfalls. Examples of the use of color coding in military and commercial cockpits are also presented.

WHEN TO USE COLOR: The advantages of using color have been demonstrated to be strongest under the following conditions:

- o When high density information is displayed in a random format (Luder and Barber, 1984; DeMars, 1975; Waruszewski, 1981).
- o When the operator must search the display for relevant information (Luder and Barber, 1984; Krebs et al., 1978).
- o When the information to be coded is qualitative rather than quantitative in nature (DeMars, 1975). That is, color is better adapted to representing conceptual information (e.g., the concept of danger) than numerical information (e.g., red = 3).
- o When attention must be drawn away from other visual tasks (DeMars, 1975; Waruszewski, 1981; Krebs et al., 1978).
- o When symbol legibility may be degraded (Krebs et al., 1978), or when different zones of information (e.g., map contours of boundaries) are difficult to distinguish solely on the basis of luminance (shades of gray).
- o When color is logically related to the operator's work (Krebs et al., 1978).

NUMBER OF COLORS TO USE: Almost everyone agrees that a color coding system for an illuminated display should use as few colors as possible and no more than four to six colors (e.g., DeMars, 1975; and Waruszewski, 1981).

COLOR AS REDUNDANT, PARTIALLY REDUNDANT, AND NONREDUNDANT CODING: The literature suggests that it is best to use color as either a redundant or partially redundant code (Krebs, et al., 1978; and Luder and Barber, 1984. Krebs et al. (1978) offer the following as an aid to deciding whether to use color as a redundant, partially redundant, or nonredundant coding dimension:

Use color as a redundant code when:

- o It is necessary to increase the detectability of a signal.
- o It is necessary to enhance an operator's ability to discriminate among many symbols.
- o It is necessary to provide a backup or alternative method of obtaining information (e.g., during degraded modes).
- o Potential users are not carefully screened for color vision deficiencies.

Use color as a partially redundant code when:

- o Information can be categorized at more than one level of specificity.
- o It is necessary to convey specific information that cannot otherwise be conveyed.

Use color as a nonredundant code when:

- o It is necessary to increase the amount of information to be displayed beyond that possible with other methods of coding.

ASSIGNING MEANING TO COLORS

What Colors to Use: Only one set of predefined color assignment schemes have been widely agreed upon. We found no other color assignment scheme suggested in the

literature. This scheme, is stated in Section 5.2.2.1.18 of MIL-STD 1472C (p. 30) and is also described and recommended by numerous other sources (e.g., MIL-HDBK-759; MIL-STD-411D; Krebs et al., 1978; and Waruszewski, 1981.) MIL-STD-1472C also recommends sizes for color coded indicator lights and relates them to coded meaning.

In a recent circular, the Society of Automotive Engineers presented a tentative list of color standards for symbology (J.D. Myers, Jr., personal communication, September 9, 1985). The recommended colors and their functional meaning are presented in Table 1.

Table 1
COLORS AND THEIR FUNCTIONAL MEANING
RECOMMENDED FOR FUTURE ELECTRONIC FLIGHT INSTRUMENT SYSTEMS

COLOR	MEANING
Red	Warnings, limits, precip 12-50 mm/hr.
Amber/Yellow	Cautions, abnormal sources, precip 4-12 mm/hr.
Tan/Brown	Earth
White	Fixed scales, fixed symbols, current values, armed modes, current route (alternate color), extreme turbulence. Current values should be separated from scale markings by intensity, character size and line width.
Green	Selected data or values, engaged modes, precip up to 4 mm/hr.
Cyan	Fixed captions, sky, armed modes (alternate). Selected data or values (alternate).
Magenta	Computed command values, analog raw data pointers, current route, precip above 50 mm/hr, extreme turbulence (alternate).

*In addition, *MIL-STD 1472C recommends* the use of Flashing Red to denote emergency conditions.

Color Consistency: It is important that the color coding scheme applied within a cockpit be consistent throughout the cockpit (Krebs et al., 1978). This principle should also be applied to systems/displays that are used in more than one aircraft.

Where to Place Colors: The ability to perceive colors is not uniform across all angles of view. Color sensitivity is greatest near the center of the fovea and deteriorates for retinal areas further away from the fovea. Additionally, the deterioration in sensitivity differs depending on the color used. The implications of this for coding have been summarized as follows (Krebs et al., 1978; and DeMars, 1975):

- o Use white in peripheral areas (i.e., areas that are outside the pilots scan and therefore will not fall within his foveal vision).
- o Red and green displays should be used only within the pilots scan (i.e., they will fall within foveal vision).
- o Place master caution and master warning displays near the center of the cockpit.

Alphanumerics and Color: Krebs et al. (1978) recommended that when alphanumerics are coded with color they should be at least 21 minutes of visual arc high and that the lines of the letters (or any color coded graphic) subtend at least 3 to 4 minutes of arc. They further state that alphanumerics should not be coded in blue. In fact they state that blue is a poor choice for coding any small object because it is not a bright color and is easily washed out. They recommend the use of red, green, and yellow when alphanumerics are to be color coded. DeMars (1975) reported that when the size of the colored symbol is continually decreased, color loss occurs in the following order: blue, red, then green.

COLOR CODING IN COCKPITS

This section discusses the use of color in existing aircraft. Primary focus is on the following military aircraft: F-16A, F-15A, F-18, F-4C, FB-111A, B-1A, B-52H, and E-2C. These were selected because they are currently operational and because data on these aircraft were available at Calspan. We also have included some information regarding state-of-the-art commercial aircraft including the Boeing 757/767, the Cessna Citation 2, and the Gulfstream 4.

It is important to note that this section describes colors that have been used in existing aircraft. It does not imply they are the best, or even appropriate, colors for the application. They are included here only because they may reflect the existence of population stereotypes. Bear in mind that since most applications of color in military aircraft have been on round-dial displays, the extent to which experience with these displays can be expected to interfere with new CRT formats is uncertain. However, when these displays are represented on the CRT using a similar or identical format as their round dial predescessor pre-existing population stereotypes should be considered.

For military aircraft such as the F-4, F-15, E-2C, and F-16, the color coding used on cockpit instruments represents primarily operational ranges or limits. In general, a green or white zone segment on the perimeter of a round dial instrument designates normal or allowable operations. Yellow zones characterize cautionary areas either above or below normal values. For warnings, the visual stimulus is almost without exception a small red line or rectangle placed over the applicable limiting reading.

The following subsections describe some dimensions in which color has been applied as a code.

Importance: Most of the current applications of color coding in cockpits has been for encoding the dimension of importance. Many of the round-dial instruments include red, yellow, or green lines/shaded areas marking off portions of the dial that represent dangerous, marginally safe, or safe values respectively. The use of red, yellow, and green indicator lights applying the standard meanings are too numerous to enumerate. In virtually all current military aircraft, red indicator lights are used to indicate warnings, yellow to indicate caution, and green to indicate status ok conditions. The Boeing 757/767 commercial aircraft deviates from this slightly in that yellow is used for displaying cautionary messages as well as advisory (i.e., status ok) messages.

Position/Direction (e.g., up, down, left, right, bank angle, pitch angle).

Vertical Reference: Although there has been no standard color convention for encoding this dimension the practice of using blue to represent sky (i.e., up) and brown to represent earth (i.e., down) is in common usage on many Attitude Director Indicators (ADIs). This is the case, for example, in the F-15 and F-16 aircraft. This code has also been adopted for the new Boeing 757/767 cockpits. In these commercial aircraft, the ADI is displayed on the CRT but is representative of the electromechanical versions. In all other cockpits reviewed, this display is not in color. Nevertheless in these other cockpits (e.g., the B-1A, E-2C, and F-4C) the top portion of this display (i.e., sky) is represented as light area and the bottom portion (i.e., ground) as a dark area.

Angle of Attack: One of the least consistent uses of color in cockpits is on the angle-of-attack indexer. This device consists of upper and lower chevrons separated by a circular symbol (donut). When the aircraft is at the optimum angle of attack, the donut is illuminated. When the aircraft is flying too slow/high-angle-of-attack, the upper chevron is illuminated; and for very fast speeds/low-angle-of-attack the lower chevron is illuminated. Table 2 contrasts the color codes used for this instrument on the F-4C,

F-18, E-2C, and B-1A aircraft. It can be seen from this table that color coding on this display has not been consistent.

Table 2
COMPARISON OF COLOR CODES FOR ANGLE OF ATTACK (AOA) ON OPERATIONAL MILITARY AIRCRAFT

AOA INDICATION	AIRCRAFT			
	F-4C	F-18A	B-1A	E-2C
	RED	GREEN	RED	GREEN
	RED	YELLOW	GREEN	AMBER
	RED	RED	YELLOW	RED

Glideslope Indicator: Currently glideslope indications are presented in two displays. The first is the ILS (Instrument Landing System) display and the second is the VASI (Visual Aid Slope Indicator). The ILS displays currently in use on the aircraft reviewed do not incorporate color. The VASI system is ground based and viewed through the windscreen during approach. It uses a nonredundant color coding system. That is, the *information regarding the aircraft's position relative to glideslope is* only displayed in color. Two red lights indicate that glideslope is too low. Two white lights indicate that glideslope is too high. One red and one white light (red on top) indicates that the aircraft is on correct glideslope. This indication actually appears pinkish until close to actual touchdown. The VASI is used on many Air Force bases and an increasing number of civilian airfields. The population stereotype associated with this color coding scheme should be considered to be quite substantial.

It is interesting to note the differences between the color codes used on the VASI system and those used in the AOA indicators. While the information coded is related, the color codes used are quite different, and in the case of the B-1A, almost opposite.

Tactical Information: Multipurpose color displays programmed for cockpits of fighter aircraft are expected to display threat data utilizing the Joint Tactical Information Distribution System (JTDIS). Threat information could be shared among many participating aircraft. While this system is still in planning, both symbol shapes and color coding have already been assigned (Ethell, 1984).

Table 3
THREAT INFORMATION CODING SYSTEM FOR THE JTIDS

Tactical Information	Shape Code	Color Code
Hostile aircraft	Triangles	Red
Unknown aircraft	Rectangles	Yellow
Friendly aircraft	Circles	Green

In addition to the color/shape coding scheme for hostile, unknown, and friendly aircraft, the JTIDS system will also display heading vectors. These heading vectors will be coded using the same color as for the associated aircraft. Navigational routing lines are displayed on the proposed tactical displays as solid yellow lines drawn between waypoints. The Forward Edge of the Battle Area (FEBA) is shown as a broken yellow line. All alphanumeric legends are displayed in white.

Early tests have identified a problem with this coding system. Pilots tended to focus on the red symbols to the almost total exclusion of the yellow symbols (Ethell, 1984). They ignored potential threats which were designated as unknowns and essentially treated them as if they were friendlies. This problem caused tactical advantage to be lost unnecessarily, since some of the unknowns were later reclassified as hostiles.

Temperature: The dimension of temperature on current military aircraft has been displayed using round dial displays. The colors used do not necessarily relate to the dimension of temperature. Rather they correspond to that aspect of temperature which relates to safety usually applying the standard red-yellow-green code described earlier. However, since the most dangerous situations are quite often associated with higher temperatures, red has usually been associated with very high temperatures, yellow with high temperatures, and green with normal or cool temperatures. In at least one instance the color blue was used to represent the lowest end of the temperature scale.

Environmental Conditions: No current use of color to encode environmental (e.g., weather) parameters has been found in military aircraft. Color is being used to code weather information on modern commercial aircraft. Notable among these are the Boeing 757/767, the Cessna Citation 2, and the Gulfstream 4. The weather radar presentation on each of these aircraft uses the following code for presenting weather:

- Red -- Heavy precipitation
- Yellow -- Moderate precipitation
- Green -- Light precipitation

This code is consistent with the general guideline of using red, yellow, and green to encode warning, caution, and normal situations, respectively.

REFERENCES

Beilman, J.L., Funke, D.J., Gawron, V.J., Hammill, H.B., Lloyd, M.M., and Mitchell, J.F. (1985, September). Guidelines for Using Airborne Color Displays. Buffalo, New York: Arvin/Calspan Corporation.

DeMars, S.A. (1975, January 5). Human Factors Considerations for the Use of Color in Display Systems. (NASA-TM-X-72196). Cocoa Beach, Florida: National Aeronautics and Space Administration.

Ethell, J.L. (1984, January). JTIDS Dividing Lines. Aerospace America, 26-33.

Krebs, M.J., Wolf, J.D., and Sandvig, J.H. (1978, October). Color Display Design Guide. (Report ONR-CR213-136-2F). Minneapolis, Minnesota: Honeywell Systems and Research Center.

Luder, C.B., and Barber, P.J. (1984). Redundent Color coding on Airborne CRT Displays. Human Factors, 26(1), 19-32.

Waruszewski, H.L. (1981, August). Color CRT Displays for the Cockpit. In AGARD Conference Proceedings No. 312: The Impact of New Guidance and Control Systems on Military Aircraft Cockpit Design.

Where the Men Are Men: Sex-Role Perceptions of Male and Female Cadets at the Air Force Academy [1]

William H. Cummings
Headquarters, Air Training Command
Randolph Air Force Base, Texas

Abstract

Studies of the personality characteristics and changes of service academy cadets have not yet examined the cadets' sex-role perceptions. This study used the Bem Sex-Role Inventory to examine male and female cadets' self-perceptions and perceptions by independent raters at the Air Force Academy. Female cadets rated themselves as significantly more feminine than their male counterparts, but not a great deal less masculine. In general, cadets saw themselves and their classmates as becoming more feminine over their four years at the Academy. Finally, both male and female cadets underrated the femininity of their same-sex classmates. This may be due to the difficulty cadets find in relating their "softer" traits to each other, particularly to their same-sex classmates.

Introduction

Among the most fascinating arenas for personality research are military training environments, and military academies may be the most interesting of these. The challenge of four years of stress, confinement, hard work, and the subsequent recognition and reward will almost certainly have interesting effects on the cadets' personalities and self-perceptions. Moreover, the introduction of women to an environment that is very traditional for men and very nontraditional for women poses further interesting questions. What happens to women during their stay at an academy? Do they become inordinately masculine (as some stereotypes would have us believe)? Do they pursue the "feminine" role all the more vigorously, in order to compensate? Or, perhaps, do the academies only attract very masculine women to begin with?

Prior studies of women attending service academies have concentrated primarily upon attitudes toward women at the academies or upon sex-related topics (e.g., Bridges, 1984). On the other hand, previous investigations of the personality characteristics of academy cadets (e.g., Hughes, 1982) have looked at both male and female cadets but have not studied sex-role perceptions. In this study, we used the Bem Sex-Role Inventory (Bem, 1974) to look at cadets' self-perceptions and the changes they might undergo during their four years at the Air Force Academy. The Bem Sex-Role Inventory (BSRI) asks subjects to rate themselves on a series of descriptive adjectives, some

---

1. The views expressed in this article are those of the author and do not reflect the official policy or position of the United States Air Force, the Department of Defense, or the United States Government.

of which are coded "masculine" ("ambitious," "self-reliant," "willing to take risks," etc.) and other adjectives coded "feminine" ("affectionate," "cheerful," "understanding," etc.) In this way, we were able to determine how masculine, feminine, or adrogynous the cadets considered themselves to be. Going one step further, we looked not only at cadets' self-perceptions (as is standard with the BSRI), but at other cadets' perceptions of our target subjects as well. This approach gave us a much more fully-rounded picture of the cadets' sex-role development. The objectives of this study, then, were entirely exploratory: What are male and female cadets' perceptions of their sex roles; how do these perceptions compare with "objective" evaluations made by other cadets; and how do these perceptions change over four years at the Academy?

Method

Subjects. Four-hundred eighty target subjects were sampled by stratified random sampling from a cadet alpha roster, with equal numbers of subjects from each sex group, class year group, and military squadron. This technique provided a target sample of 60 subjects per sex/class year group. Two-hundred twenty usable data forms were returned (45.8%), with the lowest return rate of 24 (40%) among male juniors, and the highest return rate of 34 (57%) among male freshmen. In addition, each target subject was asked to deliver a rating packet to one male and one female classmate to provide independent ratings. The return rate for male raters was 176 (36.7%) and 162 (33.%) for female raters.

The Bem Sex-Role Inventory. The Bem Sex-Role Inventory (BSRI) is a self-descriptive adjective rating scale. It asks subjects to rate themselves on a seven-point Likert-type scale ("never or almost never true" to "always or almost always true") for each of 60 adjectives. Twenty "masculine" items, 20 "feminine" items, and 20 neutral items (to prevent subjects from guessing the true purpose of the inventory) are scrambled in order of presentation, for a total of 60 items. Completion time, with instructions, is about 10 minutes. Bem determined "masculine" and "feminine" items for the inventory by discriminant analysis of her original subject pool. Bem likewise provided norms for the inventory, based on her pool of college students, such that a score of zero is neutral (neither masculine nor feminine); negative scores are masculine; and positive scores are feminine. For a full analysis, then, the BSRI provides three scores for each subject: Combined Scale (as described above, range of -13.93 to +13.93), Masculinity Scale (range of 20 to 140), and Femininity Scale (range of 20 to 140).

Method. Rating packages were mailed to the target subjects during both semesters of the 1983-84 academic year. Subjects were asked for their voluntary participation to (a) complete the BSRI (identified as the "Personal Characteristics Checklist") on themselves, and (b) give enclosed rating packages to the male cadet and the female cadet in their squadron who knew them best. The instructions to these independent raters (male and female classmates) asked for their voluntary participation in evaluating the target subject (the person from whom they received the rating packet) on the BSRI. All answer sheets were identified by code number, rather than by name, to reduce evaluation apprehension by target subjects and independent raters. (Note: This survey was simultaneously conducted at a local liberal arts

college for comparison purposes. However, space limitations do not allow discussion of these results.)

Results and Discussion

The primary analysis was a 2 X 4 X 3 analysis of variance, with Target Subject Sex (male or female), Class Year (freshman, sophomore, junior, or senior), and Rater Type (self-rater, male rater, or female rater) as the factors. Not surprisingly, male cadets were rated as significantly more masculine than female cadets ($F = 75.15$, $df = 1,534$, $p < .001$; see Table 1). This finding, essentially, simply validates the BSRI. Note two interesting related findings, though. First, both male and female cadets rated themselves on the masculine side of the scale (Table 1, Column 1). In fact, all groups of raters described both male and female cadets as neutral to strongly masculine; there were no ratings of "strongly feminine." Second, the Target Subject Sex effect was much more powerful with the Femininity scale ($F = 59.25$, $p < .001$) but occurred also to a lesser degree with the Masculinity scale ($F = 19.07$, $p < .001$; Table 2). Thus, in the eyes of both male and female cadets, and including self-ratings, female cadets are somewhat less masculine than male cadets, but a great deal more feminine.

Table 1

BSRI Combined Scores for Academy Cadets
(Positive Values = "Feminine; Negative Values = "Masculine")

	Rater Type		
	Self-Raters	Male Raters	Female Raters
Male Target Subjects	-1.96	-2.40	-1.72
Female Target Subjects	-0.22	0.08	-0.98

Of somewhat greater interest was the significant Class Year effect ($F = 2.80$, $df = 3, 534$, $p < .05$). Overall, the cadets showed a definite trend toward increased femininity during their tour at the Academy ($M = -1.26$ for freshmen; $M = -1.60$ for sophomores; $M = -0.78$ for juniors, and $M = -0.91$ for seniors). The change seems to be greatest between the sophomore and junior years ($t = 3.91$, $p < .05$, by Tukey's HSD test). Like the Target Subject Sex effect, the Class Year effect was concentrated in the Femininity Scale ($F = 3.41$, $df = 3, 534$; $p < .02$), rather than the Masculinity Scale ($F = 0.79$). Thus, it is accurate to say that the cadets tend to become more feminine over their four years at the Academy, rather than less masculine.

We can speculate on at least a few reasons for this trend: Cadets are allowed to own cars after their sophomore year, and this is usually the "green light" for the development of hitherto-neglected social graces. This interpretation is supported by our finding that the change over time was concentrated in the Femininity Scale, rather than in the Masculinity Scale:

Cadets don't become less masculine; they become more feminine. (Or, less objectively, they learn to temper their overly-masculine "macho" traits with some socially-oriented, "softer" traits.) Another big change cadets make after their sophomore year is to take squadron leadership positions. They move away from narrowly focused military training functions to more general organizational management functions. As such, they would be likely -- from sheer necessity -- to develop qualities such as "loyal," "sympathetic," and "sensitive to the needs of others," all of which are found in the Femininity scale. However, we should note that these changes are somewhat limited. The significant Class Year trend is not found among self-raters alone (F = 1.20, not significant) nor among independent raters (F = 1.67, not significant), but is supported to some extent by all three -- self-raters, male classmates, and female classmates.

To the extent that we can avoid attempts to prove the null hypothesis, the nonsignificant effects of the overall analysis are also interesting. The type of rater had no significant effect (F = 0.79), indicating that neither male nor female raters showed any strong rating bias. Furthermore, cadets' self-perceptions were generally consistent with the perceptions of their male and female classmates. Likewise, the Target Subject Sex X Class Year interaction was nonsignificant (F = 2.04, df = 3,534), indicating that the tendency toward grater femininity over time at the Academy applied equally to male and female cadets. Finally, the Rater Type X Class Year interaction (F = 1.01) and the Target Subject Sex X Class Year X Rater Type interaction (F = 0.87) were nonsignificant.

Perhaps the most interesting -- and surprising -- finding was a strong Target Subject Sex X Rater Type interaction (F = 7.23, df = 2, 534, p <.001; see Table 1). Essentially, male raters saw a tremendous difference between male and female cadets: The men are extremely masculine and the women are slightly feminine. But female raters saw much less difference: Males are somewhat masculine, but then, so are the females. As with the other findings, this effect was concentrated in the Femininity Scale (Target Sex X Rater Sex interaction, F = 6.75, p < .001; see Table 2, right half), rather than in the Masculinity Scale (F = 1.36, not significant; Table 2, left half).

Table 2

BSRI Masculinity and Femininity Scores for Academy Cadets

	MASCULINITY SCALE			*	FEMININITY SCALE		
	Rater Type			*	Rater Type		
	Self-Raters	Male Raters	Female Raters	*	Self-Raters	Male Raters	Female Raters
$\underline{M}$ Target $\underline{Ss}$	106.79	103.72	101.97	*	89.89	83.08	87.18
$\underline{F}$ Target $\underline{Ss}$	101.48	96.20	99.00	*	99.61	96.86	90.53

This is a curious finding. Given that female raters tend to blur the difference between male and female cadets, we might have anticipated that the female raters would rate their female classmates high in masculine traits. For example, the female raters might see other women at the Academy as "forceful," "athletic," and "having leadership abilities" -- to emphasize that women have enough of the "right stuff" to succeed at the Academy. Not so! Instead, the female raters underrated their female classmates in terms of femininity, whereas the male raters rated the Academy women high in femininity! So if the stereotype of female cadets is one of "unfeminine," this stereotype seems to be echoed by the female cadets, themselves. One reason for this result may be found in yet another analysis: Both male and female raters scored same-sex cadets as low in Femininity, but opposite-sex cadets as much higher in femininity (Table 2). For example, male target subjects and female raters see the male cadets as fairly high in femininity. The male raters disagree, saying that the men are low in femininity. Female raters say the same thing about female targets. This suggests that the military environment of the Academy may, to some extent, create a situation in which it is very difficult to communicate one's more "feminine" characteristics to one's classmates, especially same-sex classmates. After all, these are the kinds of characteristics that just might show that you have the "wrong stuff" or are "too soft" to succeed at the Academy.

Acknowledgments

I would like to thank four individuals who provided invaluable and greatly appreciated assistance with this study: 2nd Lt Timothy M. Ray, for his extensive data collection efforts; Capt Chris Antons, for assistance with the data analyses; and Lt Cols Richard L. Hughes and Robert A. Gregory, for their advice and consultation.

References

Bem, S.L. The measurement of psychological androgyny. Journal of Consulting and Clinical Psychology, 1974, 42, 155-162.

Bridges, A.C. Changes in attitudes toward women at the Air Force Academy. Proceedings of the Psychology in the Department of Defense, Ninth Symposium, 1984, 127-130.

Hughes, R.L. Personality characteristics of entering Air Force Academy cadets. Proceedings of the Psychology in the Department of Defense, Eighth Symposium, 1982, 420-424.

Gender-Related Situational Factors in Sex-Role Attitude Measurement[1]

Joel M. Savell
U.S. Army Research Institute
for the Behavioral and Social Sciences

Abstract

To test the effects of certain gender-related situational factors on soldier response to sex-role attitude measures and to examine the generality of published data showing such effects with college students, two experiments were conducted. Soldiers completed an embedded sex-role attitude scale in small class rooms in which the proportion of soldiers of the subject's own sex and the sex of the proctors were systematically varied. In both experiments egalitarianism scores were, as usual, lower for males than for females; but none of the other effects was significant.

About 15 years ago a study was reported (Shomer and Centers, 1970) indicating that male college students respond differentially to a sex-role attitude questionnaire as a function of the gender composition of the group in which the questionnaire is administered. In that experiment 214 male and female college students responded to two sets of questions, one set dealing with sex-role attitude and the other dealing with attitudes concerning various child-rearing practices, with two situational factors (sex of the other group members and sex of the proctor) systematically varied. Some groups consisted only of males, some consisted only of females, some were half-male and half-female, and some were all-male or all-female except for a single member of the opposite sex. Analysis of the data indicated, as expected, that the independent variables had affected responses to the two sets of questions differently. In the case of the sex-role attitude measures there was--in addition to the usual effect of subject sex (males less egalitarian on this topic than females)--a subject-sex x group-composition interaction: For male subjects but not female subjects, egalitarianism scores were higher in mixed-sex than in same-sex groups.[2] In the case of the child-rearing attitude measures, there were no effects at all.

The results of that experiment raise an interesting methodological question--one that has implications for the validity of certain kinds of data collected by the Army as well as by the other services (and indeed by any organization that is traditionally male). The question is this: Are male-soldier responses to questions concerning the role of women in the Army

----

[1]A somewhat different version of this paper was presented at the 1984 meeting of the Military Testing Association.

[2]As indicated above, Shomer and Centers (1970) distinguish several kinds of mixed-sex groups (e.g., groups that are half-male/half-female and groups that are all male except for a single female) but do not provide the statistics to show that these various distinctions are reliable.

affected by the physical presence of other people in the room while the questions are being responded to--either people of the respondent's own sex or people of the opposite sex?[3] Based on the results of the above study, (as well as various theoretical considerations) one might predict that when men respond to sex-role attitude questions in the presence of other men there is a tendency for these men to make their responses more traditional than would have been the case if they had been alone when they responded to these questions. One might also predict, based on the same results, that when men respond to such questions in the presence of women there is a tendency for these men to make their responses more egalitarian than would have been the case if they had responded to the questions alone. Each of these predictions is possible, although they are independent of each other, and neither is required by the other. Assume for the moment, however, that at least one of these predictions (or descriptions) is accurate and that the presence of at least some kinds of people--same sex, opposite sex, or possibly both--biases the responses male soldiers give to questions about sex-role attitudes. Under these circumstances, one would expect questionnaires administered in combat units, which are usually all male, to show more traditionality with regard to sex role attitude than questionnaires administered in other units, which frequently include women as well as men--not (or not necessarily) because combat soldiers are more traditional in this regard than other soldiers but because they are filling out the questionnaires under conditions that tend to elicit sex-role traditionalism. Similarly, one would expect questionnaires administered in other (i.e., support-type) units to show more sex-role egalitarianism than questionnaires administered in combat units--again not (or not necessarily) because support-type soldiers are more egalitarian in this respect than combat soldiers but because they are filling out the questionnaires under conditions that tend to elicit sex-role egalitarianism. If this effect is real, it seems likely that such questionnaires administered to soldiers in their "natural" settings (i.e., all-male groups for combat soldiers and mixed-sex groups for other soldiers) will misread the degree to which soldiers in these groups accept or do not accept the use of women in traditionally-male roles. The primary purpose of the experiments described here was to test the generality of the phenomenon reported by Shomer and Centers (1970) by replicating the study some ten years later in a military population.

EXPERIMENT I

Method

At a relatively large Army post in the southern part of the United States, 51 male and 56 female soldiers in basic training (Replication 1) and 51 male and 59 female soldiers in Advanced Individual Training (Replication 2) completed a questionnaire under one of two (same-sex vs mixed-sex) group-gender-composition conditions. In each replication, subjects were first assembled at the post testing center and from there were taken to one of four rooms that were being used for the experiment. In the case of the men, half were taken to a room (room 1) to which only men had been assigned, while the

---

[3]A similar question could of course be asked concerning the responses from women soldiers although, as indicated above, Shomer and Centers (1970) found the effect only with their male subjects.

rest were taken to one of two rooms (rooms 2 and 3) to each of which an approximately equal number of women had also been assigned. The case of the women was exactly the reverse: Half were taken to a room (room 4) to which only women had been assigned, while half were taken tc one of the two rooms (rooms 2 and 3) to each of which, as indicated above, an approximately equal number of men had also been assigned. In each replication, therefore, there were four experimental locations: one (room 1) in which the questionnaire was administered only to men, a second (room 4) in which the questionnaire was administered only to women, and two others (rooms 2 and 3) in each of which the questionnaire was administered to an approximately equal number of men and women. Each replication was thus a 2x2 (subject-sex x group-composition) between-subjects design, with subjects randomly assigned within sex to the various conditions. In the all-male group the questionnaire was administered by an enlisted man, and in the all-female group it was administered by an enlisted woman. In one of the two mixed-sex groups it was administered by an enlisted man, and in the other it was administered by an enlisted woman.

The primary dependent variable was the score obtained on a seven-item scale (Cronbach alpha =.76) that had been constructed to measure attitude concerning the role of women in the Army. Other dependent variables consisted of measures of such things as attitude regarding noncommissioned officers (NCOs).

Results

As indicated above, the variable of proctor sex was not combined factorially with the other two variables but was instead held constant (in the single-sex condition) or varied systematically (in the mixed-sex condition). The effect of this variable in the latter condition was examined by testing the proctor-sex x subject-sex interaction. In neither replication was the effect significant (both $ps > .05$), and the data from the two mixed-sex conditions were combined. In each replication, the subject's score on this scale was computed by summing his/her score on the seven items and computing the mean. The theoretical range was from 0 to 6, with higher scores indicating greater sex-role egalitarianism. In replication 1, mean scores for the men and women were 1.8 and 2.4 respectively; and the difference is statistically significant, $F\ (1,103)=23.78$, $p<.001$. The subject-sex x group-composition interaction was not significant ($F<1$), nor were any of the other effects (all $ps > .05$). In replication 2, mean scores for the men and women were 1.7 and 2.4 respectively; and the difference is again statistically significant, $F(1,106)=36.70$, $p<.001$. As before, the subject-sex x group-composition interaction was not significant ($F<1$), nor were any of the other effects (all $ps > .05$). The results of this first experiment thus failed to detect the phenomenon reported by Shomer and Centers (1970).

EXPERIMENT II

Method

Two years later, at the same post at which the first experiment was conducted, a second effort was made to replicate the Shomer and Centers (1970) results. Serving as subjects in the first session (Replication 1) were 100 male and 91 female soldiers taking Advanced Individual Training, and serving

as subjects in the second session (Replication 2) were 102 male and 105 female soldiers who were also taking Advanced Individual Training. At each session, subjects assembled in an auditorium where they were randomly asigned by sex to experimental conditions in accordance with the requirements of a 2x4x2 between-subjects design and then taken to one of five rooms which were being used for the experiment. Independent variables were (a) sex of the subject, (b) percentage of soldiers in the subjects' room who were of the subject's own sex (100%, 75%, 50%, and 25%), and (c) sex of the two proctors who administered the questionnaire in the subject's room (both male or both female). For each replication the proportion of male and female subjects in the various rooms was as follows: Room 1 (100% male), Room 2 (75% male and 25% female), Room 3 (50% male and 50% female), Room 4 (25% male and 75% female), and Room (100% female). The number of subjects in each room was approximately 20 in every case; and at each level of group-gender-composition, one of the two rooms was proctored by two male proctors and the other was proctored by two female proctors. Altogether there were four male and four female proctor pairs; and for the second session (Replication 2) the variable of proctor identity was systematically rotated across conditions. Again, the primary dependent variable was the score obtained on the sex-role attitude scale.

Results

During the two years following the first experiment, the method of scoring the sex-role attitude scale was revised (primarily by dropping one of the items and by developing a set of weights to be applied to the individual items), and scores were now able to range up to 31. The replication variable did not interact with any of the other variables in the design (all ps>.05), and the data for the two replications were combined. With respect to sex-role attitude, mean scores for men and women were 16.4 and 19.7 respectively, and the difference is statistically significant $F(1,382)=56.8$, $p<.001$. The subject-sex x group-composition interaction was not significant ($F<1$), nor were any of the other effects (all ps>.05). The results of this second experiment thus failed to detect the phenomenon reported by Shomer and Centers (1970).

DISCUSSION

Why did the interesting pattern reported by Shomer and Centers (1970) fail to replicate in the experiments described here? One possibility--which is consistent with some of the discussion by the authors of that report--is that gender was not as salient a factor in these experiments as they were in the earlier one. Responses to relevant questionnaire items confirmed that subjects were at least aware of the fact that there were or were not members of the opposite sex in their rooms; but most soldiers were in uniform when they came to the sessions, and it is possible that being dressed alike and in the common uniform served to reduce the attention that subjects paid to this fact. A second possibility is that the effect observed by Shomer and Centers, while reliable at the time, is no longer reliable because of historical/societal changes that have taken place with respect to attitudes toward women. Given, however, the controversial nature of some of the topics asked about in the questionnaire (e.,g., use of women in combat roles) and

the fact that male and female soldiers are generally recognized as differing on this question, it seems unlikely that historical/societal factors provide a sufficient explanation. A third possibility is that subjects in the recent experiments suffered from what Rosenberg (1965) called "evaluation apprehension" and, as a result, were rendered impervious to some of the forces usually operating in a situation of this sort. Efforts were made in the experiments to reduce the unnaturalness of what the subjects were asked to do, but just how successful these efforts were is not known. A fourth possibility, which requires some assumptions about the different populations represented in these studies, is that the kinds of people (soldiers? non-college young people?) who took part in the recent experiments are for some reason less susceptible to group-gender-composition effects than the kinds of people (non-soldiers? college students?) who took part in the original study. This seems unlikely, but we have no data on it one way or the other. A fifth possibility focusses on the different measures of sex-role attitude used in the studies and suggests that the measures used in the recent experiments are less affected by normative considerations than the measures used in the original study. Among the items used in the recent experiments, however, are several that are similar in character to those used in the original study; and none of these items showed group-composition effects. A sixth possibility is simply that there is no effect to replicate--i.e., that the original effect was not reliable. In this connection, it should be noted (a) that the effect reported by these authors was different in pattern from the one they originally predicted and (b) that the statistical significance of the reported effect was not great, $F(2,188)=3.39$, $.05>p>.01$.

Looking back over the results of these two experiments, it is difficult to avoid concluding that--at least for soldiers (or soldiers in uniform)--the effect reported by Shomer and Centers (1970) is not very reliable and, if it exists at all, is of limited scope. (We have not yet tried this experiment with soldiers in combat units although we have used soldiers who were undergoing basic training, and many of these soldiers were expecting eventually to be assigned to combat units.) For the time being--and until additional research on the topic suggests otherwise--there seems little reason to be concerned that surveys of soldier attitudes regarding the role of women in the Army will be invalidated by the kind of group-composition factors discussed here.

REFERENCES

Rosenberg, M. J. (1965). When dissonance fails: On eliminating evaluation apprehension from attitude measurement. Journal of Personality and Social Psychology, 1, 28-42.

Shomer, R. W., & Centers, R. (19709). Differences in attitudinal responses under conditions of implicitly manipulated group salience. Journal of Personality and Social Psychology, 15, 125-132.

Two Role Models for Predicting Decision Styles in Subordinates

Arthur B. Sweney
Wichita State University

Abstract

Instruments used to measure the Response to Power Model (RPM) and Role Reaction Theory (RRT), have been used by military organizations for research and training. The current studies report how some of these instruments relate to self-described styles on the Decision Profile (DP). Each of the six roles from the RPM (Authoritarian, Equalitarian, Permissive, Rebel, Critic, and Ingratiator) showed distinct decision patterns on the DP when the two instruments were correlated over a sample of 418 civilian workers. The five RRT subordinate pressure roles and five perceptions of the RRT superordinate roles measured on a different sample of 875 civilians showed similar and predictable decision patterns. The distinction between the Counteractive and Interactive components of the RRT was clarified by significant differences in decision patterns.

Introduction

The leadership literature has been slow in moving from the personal-trait approach to models involving roles. The role approach was introduced by Graen and Cashman (1975) with their Vertical Dyadic Linkages (VDL), by A. B. Sweney (1970, 1971), and by Sweney and Fiechtner (1976) with the Response to Power Model (RPM). These models have found mixed results in explaining the behavioral variance within dyads. Sweney (1976) found distinctly different dyadic role complementations within two sister Titan Missile Squadrons using the RPM and Graen's work seems to be difficult to replicate at lower levels within organizations, according to Vecchio (1982) and Schriesheim (1979).

The past difficulties with analyzing data across dyads may be related to the fact that both the VDL and the RPM methods involved average role behaviors rather than individual-oriented roles. Thus anticipating suggestions for individuation by Dansereou & Dumas (1977); Graen & Cashman (1975), Sweney & Fiechtner (1973) expanded the RPM model by partitioning the partner-sensitive roles (Authoritarian, Permissive, Rebel and Ingratiator) into "Interactive" and "Counteractive" components to capture whether the partners were matching or trying to change each other's role. Preliminary tests were developed by the Role Reaction Model to measure general role preferences, and individual-partner-oriented pressures and perceptions of both superordinate and subordinates. These tests have shown robust relationships in the hypothesized direction in both military (Sweney and Sweney, 1980) and civilian samples (V.A. Sweney, 1986) indicating that the core of organizational relationships has now been tapped.

The nature of the subordinate's participation in the decision making process is related to his own roles as well as those he perceives his superordinate is utilizing. Nightingale (1981) examined the subordinates reactions to the leader's style and found the greatest decision optimism and least indecisiveness under the conditions of a participative structure and a participative leader. He also identified the most pessimistic and indecisive subordinates with a nonparticipative structure under a nonparticipative leader. Tjosvold (1982) studied the effects of controversy between superiors and subordinates in leader decisions and found that the integrative position of the leader was more productive than that of either the adversarial or conflict avoidance positions. These conditions fit closely to the Equalitarian, Authoritarian and Permissive leadership roles and the Critic, Rebel and Ingratiator member roles from the RPM.

This study tested the Decision Styles of 1293 civilian employees using the Decision Profile (DP). 418 subjects were also given the Response to Power Measure and the other 875 were given two tests from the Role Reaction Theory (RRT) series. One of these, the Supervisor's Purpose Rating measures the subordinates perception of his/her leader and the other test, the Superior Subordinate Reaction Test (Subordinate Form), measures the subordinates role pressures when dealing with that particular leader. The samples were randomly selected from a wide range of industries and businesses with slightly more males than females, 264 in the first sample and 503 in the second (RRT). The average ages of the respondents were 34.7 years for the RPM sample and 33.9 for the RRT sample.

The tests were scored and Pearson product-movement correlations were calculated between the thirteen Decision Profile scores and the six RPM scores or the five SSRT and five SPR scores depending upon the sample being analyzed. The significance levels were established using the more conservative two-tailed test since the exploratory nature of the study precluded making firm hypotheses. Only those correlations significant at the .01 level were reported.

Results

The significant correlations between the decision scales and the roles as measured by the RPM are found in Table 1. Thirty-four of the seventy eight possible combinations were significant beyond the .01 level. Permissives and Ingratiators were found to be Risk Avoidant while Rebels were Risk Seekers. Permissives, Rebels and Ingratiators displayed Post-decision Dissonance while Critics did not. The Internal focus of control posited by Rotter (1966) seems to apply to both scales 5 and 8 where the Equalitarian and the Critic express high control and optimism while the four partner-oriented roles show pessimism and loss of control over outcomes. The Permissives and Rebels show the most dysfunctional decision patterns and the objective Critics show the most favorable ones.

The correlations between the decision scales and the subordinate roles from the RRT model are found in Table 2. Over one-half of the scale combinations were significant beyond the .01 level. The same general patterns for the subordinate roles were shown with some notable exceptions. The resistant Rebels (R_C) showed anxious Wildness while the demanding Rebels (R_I) did not. The dependent Ingratiators (I_I) were much more Risk Avoidant then were the

Table 1: Correlations between
Organizational Preference Roles (RPM) and Decision Styles (DP)
N = 418

	Organizational Role Preferences - RPM					
D.P. Styles	Author.	Equal.	Permissive	Rebel	Critic	Ingrat.
Risk Avoidance	-	+	315*	-118	+	216*
Post Decision Dissonance	+	-	339*	170*	-215*	132
Dependent Indecision	+	-	309*	+	-	161*
Rigid Indecision	+	-	377*	+	-146	143
Lucky Expectations	+	146	-	-165*	244*	+
Fast Decisions	+	+	-193	-	+	-
Luck Dependent	+	-	+	252*	-128	-
Uncontrolled Outcomes	+	-	278*	223*	-184*	123
Hopeful Preservation	+	127	-	-	258*	+
Conservative Shifts	119	-	+	114	-	-
High Risk/High Gain	125	-126	-	142	-	-118
Anxious Wildness	+	-	+	258*	-217*	-
Conscious Processing	-	+	-	-282*	-238*	+

Only significant (.01) correlations have been given. Non-significant correlations are identified only by direction. Decimal points have been dropped.
* = .001 level of significance

Table 2: Correlations between
Subordinate Role Pressures (SSRT-sub) and Decision Styles (DP)
N = 875

	Subordinate Role Pressure (SSRT)				
Subordinate Decision styles	Rebel Counter (Resistant)	Rebel Inter (Demanding)	Critic (Objective)	Ingrat. Inter (Dependent)	Ingrat. Counter (Supportive)
Risk Avoidance	-162*	-214*	-	206*	+
Post Decision Dissonance	+	-092	-202*	113	094
Dependent Indecision	+	-	-111	+	+
Rigid Indecision	+	-126*	-208*	196*	153*
Lucky Expectations	-093	-	+	+	+
Fast Decisions	124*	154*	-116*	-216*	-117
Luck Dependence	+	+	-333*	-	+
Uncontrolled Outcomes	+	-121*	-331*	246*	222*
Hopeful Perseveration	-	+	225*	-	-116*
Conservative Shifts	+	-	-180*	+	+
High Risk/High Gain	+	123*	+	-141*	-
Anxious Wildness	147*	-	-178*	+	+
Conscious Processing	-	+	218*	-128*	-232*

Only significant (.01) correlations have been given: Non significant correlations are indentified only by direction. Decimal points have been dropped.
* = .001 level of significance.

supportive Ingratiators (I_C). The demanding Rebels (R_I) felt greater Controlled Outcomes than did the resistant Rebels (R_C). In these and other respects the demanding Rebel (R_I) showed patterns much closer to the Objective Critic than did any of the other roles.

The effect on decision patterns of the subordinates perception of his/her leader is shown in Table 3. About one-third of these correlations were significant at the .01 level. Subordinates with directive leaders (A_I) were pessimistic, have not compensated for their uncertainty by being slow, Skill Dependent, keeping Control over Outcomes, utilizing Secure Discipline starting out conservative and Shifting Risky with experience and depending on Conscious Processing. Subordinates who saw their leaders as Equalitarian showed Firm and Flexible Decisiveness, had Lucky Expectations, and utilized Secure Discipline and Conscious Processing. Those who saw these leaders as Restrictive (A_C) showed Luck Dependence, Uncontrolled Outcomes, Discouraged Vacillation, Anxious Wildness, and responded to Intuitive Hunches.

Table 3: Correlations between
Perceptions of Superordinates (SPR) and Subordinate's Decision Styles (DP)
N = 875

	Perceptions of Superordinate's Role (SPR)				
Subordinates DecisionStyles	Author. Counter. (Restrictive)	Author. Inter. (Directive)	Equal. (Objective)	Permiss. Inter. (Compliant)	Permiss. Counter. (Protective)
Risk Avoidance	-	+	+	-113*	-
Post Decision Dissonance	+	-	-126*	+	+
Dependent Indecision	+	-	-	-	+
Rigid Indecision	+	-	-153*	+	+
Lucky Expectation	-	-142*	136*	-093	-
Fast Decisions	+	-085	-	+	+
Luck Dependence	156*	-163*	-	+	+
Uncontrolled Outcomes	147*	-215*	-	+	-
Hopeful Perseveration	-091	097	+	-	-
Conservative Shifts	+	-093	-	+	+
High Risk/High Gain	+	-	+	+	-
Anxious Wildness	087	-097	-139*	121*	113*
Conscious Processing	-164*	134*	093	-105	-

Only Significant (.01) correlations have been given. Non-significant correlations are identified only by direction. Decimal points have been dropped.
* = .001 level of significance.

Discussion

This study was one of the few which tries to operationalize the consequences of different dyadic role relationships. The results tend to confirm the general assumptions about leadership roles, but provides the added

specificity about subordinate reactions to these roles. The differences between the Interactive and Counteractive components of the roles are clarified.

References

Dansereau, F., and Durmas, M. "Pratfalls and pitfalls in drawing inferences about leader behavior in organizations." In J. Hunt & L. Larsen (Eds) Leadership: The Cutting Edge, Carbondale, Ill.: Southern Illinois University Press, 1977.

Graen, G. and Cashman, J. "A Role-making model of leadership in formal organizations." In J.G. Hunt & L.L. Larsen (Eds) Leadership Frontiers, Kent, Ohio: Kent State University Press, 1975.

Nightingale, D. "Participation in decision-making: An examination of style and structure and their effects on member outcomes." Human Relations, 1981, 34, 1119-1133.

Rotter, J.B. "Generalized expectancies for interval vs. external control of reinforcement." Psychological Monographs, 1966, 80(1) whole 609. pp. 1-28.

Schresheim, C. "The similarity individual directed and group directed leader behavior descriptions." Academy of Management Journal, 1979, 22, 345-355.

Sweney, A.B. "Organizational power roles." Professional Management Bulletin, 1970, 10, 5-13.

Sweney, A.B. Leadership: Management of Power and Obligation. Wichita, KS: Test Systems International, 1981.

Sweney, A.B. & Feichtner, LA. "Factors of subordinate-superordinate interaction", Multivariate Experimental Clinical Research, 1976, 2, 93-100.

Sweney, V.A. & Sweney, A.B. "Validating the role reaction model for predicting superior-subordinate relationships." Preceedings: Psychology in the Department of Defense, Seventh Annual Symposium. Colorado Springs: USAF Academy, 1980.

Tjosvold, D. "Effects of approach to controversy on superior's uncorporation of subordinate's information in decision making." Journal of Applied Psychology, 1982, 67, 189-193.

Vecchio, R. "A further test of leadership effects due to between-groups variation and within-group variations." Journal of Applied Psychology, 1982. 67, 200-208.

Aerospace Clinical Psychology (ACP)

Timothy S. Strongin, Ph.D., Capt, USAF[1,2]
United States Air Force School of Aerospace Medicine

Abstract

ACP is a growing subspecialty of military psychology. Throughout the history of military aviation, issues of fear and anxiety have emerged as critical to mission accomplishment. During peacetime, ACP focuses on performance enhancement, safety, and retention efforts. Existing psychological techniques are applied to unique problems in ACP. Neuropsychological evaluation of aircrewmen, biobehavioral interventions, organizational consultation, demonstration programs and proposed training courses are describei.

Introduction

Aerospace Clinical Psychology is the most recent addition to the Aerospace Medicine Team. Expertise in the description, explanation, and alteration of human behavior is essential to the aeromedical mission. Performance enhancement is achieved by teaching effective stress management techniques, cognitive rehearsal and performance strategies and by developing leadership skills. Safety is improved by recognizing and changing behavior patterns that can lead to aircraft mishaps. Retention is improved by recognizing personal and familial difficulties in time to resolve them. Retention also benefits from life-style alterations associated with the Air Force's Coronary Artery Risk Evaluation program. It costs as much to replace a flier who quits or develops mild heart disease as one who dies in an accident.

---

[1]The author wishes to gratefully acknowledge the assistance and support of Dr. (Col Ret) David R. Jones, Capt (Dr) David A. Giles, and Dr. John C. Patterson of the USAF School of Aerospace Medicine for their assistance in the development of these ideas.

[2]The opinions and findings of this paper are those of the author and do not necessarily reflect those of the Department of the Air Force or the Federal Government.

Historical Development

During World War I, Anderson (1919) and Pilmore (1919) documented the importance of stress, fear, and personality factors in learning to fly and in successful combat operations. Problems with fear, fatigue, and characterological problems were noted to be relevant to both performance and safety. During the interwar years, little was written about the clinical psychology of aviation. In 1939, Armstrong noted the critical importance of life changes in disrupting the psychological adjustment to the demands of aviation.

No fliers were subjected to greater situational stress than those in WW II. Davis (1946) wrote of the importance of psychological sophistication in flight medicine. Bond (1952) and Grinker and Spiegel (1945) were explicit in their calls for centralized psychological/psychiatric treatment for fliers by specialists with particular expertise and experience in aviation. Familiarity with the fliers, their environment, and their stresses needed to be combined with psychological sophistication to prevent small problems from becoming disruptive or dangerous.

In Korea, the U.S. Air Force confronted the problem of fear of flying. Reservists refused to train in new aircraft and cited anxiety as their reason. After arriving in Korea, most fliers did exceptionally well. Gatto (1954) noted that psychologically sophisticated practitioners, acquainted with the fliers and available at each base would have greatly diminished the loss of skilled fliers to psychological dysfunction. No professional literature was published on psychological problems among fliers in Vietnam.

In the last 30 years, the Israelis have developed a program involving psychologists attached to the flying units at each base. The Luftwaffe of the Federal Republic of Germany relies on aviation psychology for clinical evaluation treatment, mishap investigation, pilot selection and research. The U.S. Navy has long enjoyed the services of aviation psychologists with training in industrial, experimental, and engineering psychology. Clinical psychologists, however, have not been formally associated with the flying mission in the U.S. Armed Forces except on a local basis. In 1985, the USAF in Europe established formal liaison between clinical psychologists and flight surgeons. Part of the program included a 40 hour course, orienting psychologists and flight surgeons to the history and applications of ACP.

Fear of Flying

The fear of flying (FOF) has long troubled those associated with aircrewmen (Jones, in press). It is unclear whether FOF refers to a specific neurosis, an adjustment disorder, a characterological weakness, or the result of fatigue. A review of the literature (Strongin 1986) suggested that evaluators should ask themselves about the underlying personality of the flier in question, the stresses to which he's been subjected, and finally, changes in

life circumstances that may have altered the flier's motivational or defensive structure.

Neuropsychological Issues

Neuropsychological testing is important in dealing with fliers. There is no doubt that g induced loss of consciousness will continue to endanger military fliers. Short of unconsciousness, what cognitive abilities suffer from sustained moderate g loading? What is the effect of years of high performance flying on subsequent cognitive function? What neuropsychological talents are required to become a flier in the first place? These questions are under study at the U.S. Air Force School of Aerospace Medicine (USAFSAM). Certainly, fliers score much higher on tests of cognitive function than the population at large.

In addition to neuropsychological questions, the clinician working with fliers needs to know whether and how fliers differ from normative populations on personality tests. Recent research (Retzlaff and Gilbertini, 1985) confirms previous research (Fine and Hartman 1968) and suggests that there may be several clusters of personality traits associated with entry into military flight training. Once qualified, aircrewmen tend to cope with problems differently than other people. They spend less time on introspection and more time on problem solving, environmental manipulation, and are usually able to exercise denial effectively. Motivation is critical to performance. It has long been noted (Armstrong, 1939, Tempereau, 1956) that motivation to fly changes with maturation. The use of psychological evaluation techniques clarifies these changes. The MMPI and Reitan-Halstead have been normed for use with fliers. A new version of the TAT, specifically for use with military fliers, is under development at USAFSAM.

Biobehavioral Intervention

A particular area of interest at USAFSAM is that of the interaction between personality and physical health. A multiyear study of fliers with asymptomatic cardiovascular problems is underway at USAFSAM. Whether definable personality types are associated with the development of heart problems will clarify whether preventive mental health intervention is needed. Avoiding losses due to medical problems will more than compensate the cost of researchers and clinicians in saved training costs.

Another aspect of the overlap between ACP and aerospace medicine is the treatment of airsickness. Jones, (et al, 1985) demonstrated the value of biofeedback training in helping fliers overcome airsickness. At USAFSAM, we are studying a variety of biofeedback measures to see which combination is most effective at interrupting the pattern of behavior culminating in airsickness. The reports of professionals in this area describe motivation as necessary but not sufficient to overcome airsickness. Those involved in ACP need to be skilled clinicians as well as technicians if they seek success with biobehavioal techniques.

Working With The Flight Surgeon

Where clinical psychologists have been attached to flight surgeons' (FS') offices, consistent reports of successful interaction have followed. At a flight training base, ACP was effective in treating numerous cases of airsickness and in using behavioral techniques to enhance learning. Many otherwise well qualified students who would have been unnecessarily washed-out were retained (Giles, 1984). At an overseas fighter base, an ACP helped nearly half of the families in one operational squadron. There were no divorces in the squadron for the 18 months of his involvement and the medical and operational leadership found the program so useful that the U.S. Air Force Europe has instituted a similar program at eight operational bases.

In the future, ACP will have a dual focus. First, training in ACP will precede attachment to FS' offices. Once on base, the ACP will work with the FS to educate flight commanders about stress and coping among fliers, and to provide knowledgeable consultation to the FS, the leadership, the fliers, and their families. The goals will be to enhance performance, improve safety, and to enhance retention. Behavioral changes that promote coronary artery risk reduction will be promoted. The second focus of ACP will be research. The expansion of hard, relevant data is necessary to make the most of ACP's consultative role.

Training

Training programs will support operational needs. First, we hope to establish an 80-240 hour course in applied ACP as a prerequisite to working with the FS. Course work would include: an overview of military operations; aeromedical support of military operations; the historical development of Aerospace Medicine and ACP; traditional aviation psychology; pilot personality; clinical evaluation techniques; consultation and liaison with the FS; combat psychology; biobehavioral intervention with fliers; brief psychotherapy; marital therapy; organizational development and consultation; ethical considerations.

The need for human factors expertise in the FS' office goes without saying. ACP offers greater breadth, depth, and experience than is presently available on the aeromedical team. The means to associate ACP with Aerospace Medicine has been demonstrated and coursework is ready for delivery. Clinical psychology makes our fliers better and safer. It reduces unnecessary losses due to personal problems, stress, illness, demotivation and coping patterns.

References

Anderson, H.G. The psychology of aviation in The Medical and Surgical Aspects of Aviation. London: Oxford University Press, 1919.

Armstrong, H.G. Principles and Practice of Aviation Medicine. Baltimore: Williams and Wilkins Company, 1939.

Bond, D.C. The Love and Fear of Flying. New York: International Universities Press, 1952.

Davis, D.B. Phobias in pilots. The Military Surgeon, Aug 1945, 105-111.

Fine, P.M., and Hartman, B.O. Psychiatric strengths and weaknesses of typical Air Force pilots. USAF School of Aerospace Medicine, TR-68-121, Brooks AFB, TX, 1968.

Gatto, L.E. Understanding the "fear of flying" syndrome. U.S. Armed Forces Medical Journal, 1954, 5, 1093-1116 and 1267-1289.

Giles, D.A. Personal communication, 1984.

Grinker, R.R., and Spiegel, J.P. Men Under Stress, Philadelphia, Blakiston, 1945.

Jones, D.R., Levy, R.A., Gardner, L., Marsh, R.W., and Patterson, J.C. Self-control of psychophysiologic response to motion stress: Using biofeedback to treat airsickness. Aviation, Space, and Environmental Medicine, Dec, 1985, 1152-1157.

Jones, D.R. Flying and danger, joy and fear. Aviation, Space, and Environmental Medicine, in press.

Pilmore, F.U. The nervous element in aviation. Naval Medical Bulletin, 1919, 13, 458-478.

Retzlaff, P.D., and Gibertini, M. Air Force pilot personality: Hard data on the "right stuff." Unpublished manuscript, 1985.

Strongin, T.S. An historical review of the fear of flying among aircrewmen. Paper submitted to Aerospace Medicine and Environmental Health, 1986.

Tempereau, C.E. Fear of flying in Korea. American Journal of Psychiatry, 1956, 113, 218-222.

A Psychological Analysis of Counterespionage

Richard W. Bloom
Air Force Intelligence Service
Washington, D.C. 20301

Abstract

Espionage is an ongoing threat to national security. This paper provides a psychological critique of five popular recommendations to counter espionage--(1) increasing the severity of penalties, (2) decreasing the number of individuals who hold security clearances, (3) decreasing the amount of information that is classified, (4) increasing the comprehensiveness and frequency of security investigations, and (5) increasing the use of the polygraph. An additional recommendation based on the psychological research on moral development is presented and critiqued. The paper concludes with a series of research topics that psychologists interested in counterespionage may wish to pursue.

Espionage is an ongoing threat to our national security ("Year of the Spy," 1985). Popular recommendations to counter espionage also may threaten security. This paper provides a psychological critique of these recommendations and a psychological alternative to them.

One recommendation is to increase the severity of penalties for convicted spies ("Keeping the Nation's Secrets," 1985, p. 50). This will only affect the real and potential spies who expect to be apprehended. Moreover, a seemingly severe penalty may be perceived quite differently by its target. A spy with ideological motives or one who "does it for the kicks" may find more severe penalties to be positively reinforcing for espionage behavior (cf. Zuckerman, 1964). In addition, real or threatened punishment often attenuates the aversiveness of later punishment, decreasing its influence on behavior (Hunt, 1979). It also may have a more unpredictable effect on perceptions and behavior than positive or negative reinforcement (Bloom, 1969). Punishment does provide us with a sense of control and satisfaction after being violated which belies our fragile grasp on national security.

A second recommendation is to reduce the number of individuals who have security clearances ("Keeping the Nation's Secrets," 1985, p. 17). This makes sense, if there are too many to begin with. The proper number of cleared individuals would depend on the amount of information that truly needs to be classified along with the nature of classified procedures. An arbitrary reduction in cleared personnel well could harm national security. Reducing the number of cleared personnel also ignores individuals without clearances who can induce espionage as case officers or agents of influence. Finally, it increases the potential value of each cleared individual who can be exploited through hostile intelligence as well as the potential damage to national security from each spy.

A third recommendation is to reduce the amount of information that is classified ("Keeping the Nation's Secrets," 1985, p.31). Again, this makes sense, only if there is too much to begin with. Otherwise, declassifying classified information would facilitate the intelligence activities of our adversaries. Note, also, that reducing the amount of classified information increases the value of open source data to our adversaries and renders our classified reservoir ever more valuable. Moreover, with everchanging developments in technology and alliances, with the contributions of accident, serendipity, and fate, the same information properly may be classified, then

unclassified, then classified again. Analogous to so much of psychopathology, reclassifying the unclassified may attenuate our anxiety for security but be dysfunctional as security behavior.

A fourth recommendation is to effect more intensive and more frequent investigations of individuals who may have, do have, and have had security clearances ("Keeping the Nation's Secrets," 1985, pp. 20; 23). This would entail a huge investment in money and personnel--both for investigations and various support activities. The value of this investment depends on how much we know about variables that predict espionage. If we know little, how will investigators know what to look for--save for clandestine meetings between unsavory characters wearing ill-fitting trenchcoats? Although money, sex, ideology, kicks, and real or imagined slights are most often admitted as rationales by convicted spies ("Why the Market," 1985), our unclassified directives, regulations, and guidelines do not delineate them explicitly. Even if they did, their literal and consistent application would result in huge, false positive errors (cf. Director of Central Intelligence Directive No. 1/14, 1981). On the other hand, there is explicit delineation of statistically deviant and socially stigmatic traits such as homosexuality and mental disorders (cf. DOD 5200.2-R/AFR 205-32, Appendix 1, p. 107). Unfortunately, there is no unclassified research or data base linking these traits with the act of espionage.

A fifth recommendation is to use the polygraph more frequently for personnel security screening ("Keeping the Nation's Secrets," 1985, pp. 25-26). Unfortunately, there is significant, scientific evidence supporting the polygraph's validity only in criminal investigations, not in personnel security screening nor in the screening of large numbers of people in investigations of unauthorized disclosure (Saxe, Dougherty, & Cross, 1985). Moreover, certain kinds of individuals either "look like" they're lying when they are not, e.g., some intropunitive neurotics, or "look like" they're not lying when they are, e.g., some psychopaths (Waid, Orne, & Wilson, 1979). In fact, some internationally recognized authorities on the polygraph have termed the very use of four or five crude psychophysiological measures to detect the complex social psychological phenomenon of deception as absurd (Lykken, 1974). Although the polygraph may "work" in any situation where the individual bestows it magical powers and confesses accordingly, more sophisticated individuals may suffer the slings and arrows of false positive errors or enjoy the heady draft of false negative ones. Further, there is some research to suggest that individuals can be trained to "beat the machine" through physical (Hodes & Honts, 1982) or psychological (Corcoran, Lewis, & Garver, 1978) countermeasures.

These five recommendations are all alleged deterrents. In conjunction with common techniques of physical and operational security they are designed to decrease the probability that individuals will commit espionage.

In a more proactive fashion, however, what can be done to decrease the probability that individuals wish or intend to commit espionage? One way would be to delineate how individuals judge right from wrong, and then develop a security management system that reinforces the act of espionage as wrong in the individual's own terms.

In fact, there is extensive, psychological research on the epigenesis of judging right from wrong--or moral development (Colby & Kohlberg, 1984). A state of the art perspective suggests that most individuals of different cultures, societies, sexes, races, and ethnicities can sequence through up to six moral stages (Snarey, 1985). Each moral stage is associated with a specific cognitive basis for judging right from wrong.

Stage 1 is termed obedience and punishment orientation. To these individuals, what is right is what avoids punishment or physical damage. Right also is being obedient for the sake of being obedient. An individual in stage 1 would be less likely to view espionage as right, as penalties for it became more severe.

Stage 2 is termed instrumental purpose and exchange. To these individuals, what is right is what is in one's immediate personal interest. Right is part of an exchange perceived as equal or a "good deal". An individual in stage 2 would be less likely to view espionage as right, as one's country provides positive reinforcers of suitable quality and quantity, viz., promotions, recognition, prestige, and money. One must receive what one perceives one deserves.

Stage 3 is termed interpersonal accord and conformity. To these individuals, what is right is what is in accord with the expectations of one's friends, colleagues, and admired others. An individual in stage 3 would be less likely to view espionage as right, as one's significant others felt similarly.

Stage 4 is termed social accord and system maintenance. To these individuals what is right is what they've agreed to do. Laws always are to be upheld, except in extreme cases where they conflict with other important and fixed social duties. Right is what contributes to the common good. An individual in stage 4 would be less likely to view espionage as right, if one has agreed not to do it or if one perceives it as not in the best interest of one's country. A possible exception could be fear for the life of loved ones.

Stage 5 is termed social contract, utility, individual rights. These individuals recognize that what is right may be relative for different individuals. They believe one usually should uphold what a majority of one's reference group considers to be right. They believe that a few values, rights, or proscriptions may be nonrelative and must always be upheld. If a proscription against espionage is not one of this small, nonrelative group, the individual may be a significant security risk.

Stage 6 is termed universal ethical principles. These people believe in a set of personally derived, but universal, moral principles. They believe one should act on these principles, even if they violate laws or social agreements. For example, if a proscription against espionage conflicted with a belief in the equality of human rights or in the dignity of human beings, it could follow that the proscription should be violated.

This research in moral development poses a complex challenge for developing a security management system. First, since there may be at least six cognitive bases for judging right from wrong, the system must provide six sets of reinforcers linking the act of espionage with right. Second, the system must ensure that the appropriate set of reinforcers is consistently applied to individuals with the corresponding cognitive bases for judging right from wrong. For example, laws and agreements may be irrelevant to stage 6 individuals. Third, the system must ensure that each set of reinforcers is, indeed, perceived as such by individuals with the six corresponding cognitive bases for judging right from wrong. What is perceived as a "good deal" by a security supervisor may not be by a stage 2 supervisee. Fourth, the system must ensure that the six sets of reinforcers do not interact in a manner detracting from their unique, individual effects on their intended targets. The harsher penalties effective with stage 1 individuals may decrease the reinforcing power of the interpersonal accord and conformity factors crucial to stage 3 individuals, much as intrinsic reinforcement often is subverted by extrinsic reinforcement (Harackiewicz, Manderlink, & Sansone, 1984). Fifth,

the psychological research strongly suggests that each individual can vary in stage of moral development through time and for different kinds of situations (Snarey, 1985). Thus all six sets of reinforcers must be maintained, because each could be operative for the same individual throughout a security-related career. A last challenge--there is rarely a perfect, often not even a highly significant correlation between cognitions and behavioral intentions on the one hand, and behavior on the other. Part of the problem is that affects and motives also influence behavior. Another is that intrapsychic factors only partially influence behavior. Biological, e.g., neuroendocrinological and genetic, sociological, e.g., cultural and economic, and environmental, e.g., topographical and meteorological, also can be behaviorally relevant. Although significant relationships between moral judgment and moral behavior have been demonstrated (Malinowski & Smith, 1985), they cannot be taken for granted.

Of course, several basic problems in psychological assessment become even more salient when espionage is of concern. People sincerely may report aspects of their own minds and bodies that they could not possibly know (Nisbett & Ross, 1980). There are consensus and self-symbolizing effects, as well as other self-serving biases, which make it difficult to establish linkages between intrapsychic phenomena and behavior (e.g., Gollwitzer & Wicklund, 1985; Wetzel & Walton, 1985). In addition, attempts to dissimulate on psychological inventories differentially affect obvious and subtle items (Dubinsky, Gamble, & Rogers, 1985). Moreover, individuals apparently differ in how they integrate information during moral judgment (Forsyth, 1985).

News accounts of U.S. citizens who publically have been charged and/or convicted of espionage bear out the frequent incongruence between what one judges to be right and what one does. Most seem to have taken a course of action they knew to be morally wrong (e.g., "Court denies bail," 1985). Most have confronted a personally intolerable disparity between what they have had and what they have wished to have, between the real and the ideal, and have broken the linkage between moral judgment and moral behavior. Unfortunately, just as moral weakness can be associated with the act of espionage, the moral development approach to espionage has its weakness.

Perhaps the crucial issue is that we are all amoral to a degree through omission and commission. This cardinal characteristic of human existence led Camus to cherish his nights of despair--only then was he truly living. Only then was he truly human. Only then could he savor the possibility of being truly moral (Camus, 1951/1980).

It should not be surprising that espionage has been termed the second oldest profession. As with its predecessor, it may always be with us. It may be better managed, but not eradicated. Thus the term, security management. Psychologists interested in contributing to security management may pursue the complexity of the psychological approach through (1) further study of moral development and its assessment, (2) delineation of situational, intrapsychic, and other variables in accordance with the biopsychosocial model which affect moral behavior, (3) assessing the fit of security management systems with the cognitive bases and other substrates used by individuals in judging right from wrong and accordingly, (4) providing clinical and counseling services to decrease the disparity between the real and ideal for each security-related individual, (5) through "dust-bowl empiricism" identifying psychological characteristics associated with the act of espionage--e.g., narcissism, immaturity, impulsivity, and private self-consciousness, and (6) providing the fruits of their developing knowledge to security investigators and adjudicators (See Personnel Security, 1984).

Through this pursuit, psychologists can challenge the anonymous security

specialist who has said, " there's just no way you can look inside somebody's brain and know what he's going to do (Why the Market, 1985).

REFERENCES

(Due to space limitations, references have been omitted from this edition of the paper. References are available upon request from the author at the address given at the beginning of the paper.)

DISCLAIMERS

This paper does not necessarily reflect the views of the Air Force Intelligence Service or the Department of Defense.

Accelerated Learning Method Based on the Lozanov Method

Carl Schleicher
Lozanov Learning Systems Inc.
Silver Spring, Md.

Abstract

In recent years a new method of accelerated learning of rote material has been developed which is known as the Lozanov Learning Method. This is also known as Suggestopedia or "Superlearning" and represents a stress-free technique to enable students to learn a large amount of material in a short period of time. The success of this method is due primarily to the elimination of blocks and barriers which inhibit the learning process. Results to date have demonstrated that through this method, learning is fast and effective, with memory retention as high as 95% of the material learned and this is over long periods of time. The method works by the use of special techniques employing the "Psychology of Suggestion" as developed by the Bulgarian Psychiatrist Dr. Georgi Lozanov. This method has been used by the Defense Language Institute, ARMCO, Ford Motor Company, and many schools and institutions of higher learning. This method represents an alternative to the type of traditional "forced learning" approach which are used in many U.S. educational institutions especially the U.S. service academies.

Background

A new method of education has emerged in the 1980's that promises to help mankind keep pace with the accelerated development of human potential capacities as well as with the information explosion phenomena caused by the audio-visual technology. This method is also of special interest in that it enables tapping of man's reserve capacities in a holistic sense whereby the mind and body are totally involved in the learning process.

This dynamic program, initially developed in Bulgaria about 1956 by Dr. Georgi Lozanov, has received wide acclaim from such diverse sources as UNESCO, U.S. Foreign Service Institute, Maurice Thorez Institute in Moscow (USSR), numerous universities including Georgetown and Howard in the District of Columbia. The method itself is described in detail in the succeeding paragraphs.

The Lozanov Learning Method, also popularly known as Suggestopedia, is a dynamic, stress-free learning technique which enables one to learn a great amount of material in a short period of time. Its success is due primarily to the elimination of barriers which inhibit the learning process. The Lozanov Method involves the student in role-playing, music, art, games, and mental relaxation to stimulate his/her creativity. When the mind is free from stress and engaged in a variety of enjoyable activities, a greater amount of information can be absorbed directly into the long-term memory with greater ease. As a result, learning is fast and effective, with memory retention as high as 95%. And, contrary to student performance under conventional teaching methods the amount of material retained and used remains the same, and often increases, over long periods of time.

The method was developed by Dr. Lozanov who is an experienced psychiatrist and psychotherapist from Sofia, Bulgaria. He is currently the Director of the

Institute of Suggestology which he helped found in 1966. Dr. Lozanov is a graduate in medicine from the University of Sofia and later received his Ph.D. from the University of Kharkov in the U.S.S.R. for his dissertation, which dealt with the positive application of suggestion in the fields of medicine and education.

At the present time the Lozanov Method is being used to teach stress-free language courses in many foreign languages including Spanish, French, Portuguese, German, English, Arabic, Chinese, and Swahili as a second language. The Lozanov Learning Systems, headquartered at 1315 Apple Avenue, Silver Spring, Maryland, U.S.A. with its specially trained teachers, is the principal organization using the Lozanov Method for language instruction in the United States.

During the class, the student will acquire a vocabulary of about 2,000 new words and be able to converse using over 1,400 words. At the same time, the student will be learning about the customs, cultures and mores associated with the language. The student will not only be learning quicker than with conventional methods, but will feel refreshed after each class period.

The basic Lozanov language courses in Latin-alphabet languages last 72-75 hours and Non-Latin alphabet languages last 90 hours due to the need to learn a new alphabet. Lozanov Learning Systems (LLSI) employs only those teachers who are not only well qualified to teach a foreign language but who are specially trained in teaching foreign languages using the Lozanov Method. LLI's highly trained teachers all speak with native fluency. Most Lozanov teachers have taught foreign language for many years using both conventional methods and the Lozanov Method.

The only requirement of a student is a desire and willingness to learn. In fact, the Lozanov Method has been highly successful in teaching foreign languages to individuals who have previously had difficulty learning a foreign language using conventional methods. Lozanov language courses are stress-free experiences in learning in which one will learn a great amount of material in a short period of time -- while at the same time increasing mental capacities. There is no rote memorization or tedious repetition of dialogues. In addition, there is NO HOMEWORK, no cramming, or pressure. Yet the student will learn a new language in a small fraction of the time it takes with conventional methods.

Description of Method

The Lozanov Learning Method (LLM) takes advantage of the modern neural-physiological finding that long-term memory is a function of the whole brain and not of any localized area within the brain. It assumes that since long-term memory is a function of the whole brain, then learning must also be a function of the whole brain. It attempts therefore to include both emotional (sub-cortex) and rational (cortex) aspects of behavior in learning process, and it attempts to integrate the verbal-intellectual-sequential mode of understanding which, in naturally right-handed people, primarily characterizes the left hemisphere of the cortex with the intuitive-holistic-spatial mode of understanding which, in naturally right-handed people, primarily characterizes the right hemisphere. In keeping with the theory that long-term memories are best implanted when the whole brain is involved in the learning process, learning activities under LLM are never undertaken in isolation from other activities that traditionally have been looked upon as irrelevant to, if not distracting to, the learning process. Whole-brain learning characterizes each of the three stages in the LLM learning sequence. The three stages in this sequence are: decoding or deciphering, concert, and elaboration. These stages are described in detail as follows:

Stage 1: Decoding

In decoding the teacher introduces key words and acts out their meanings The presentation is dramatic and, when well done, interesting. The teacher involves the students as active participants in the process by asking questions. The student's native language is never used, only the target language. Acting out words is better demonstrated than described. In addition to acting out key words, the teacher also presents paradigms and explains target language structures, always without recourse to the students' native language. The decoding stage of a typical lesson takes about 10-15 minutes.

Stage 2: Concert

The concert stage is divided into two phases. The first is called the active concert session and the second the pseudo-passive concert session. Each lasts about 45 minutes.

In the active concert session the students have the text of the dialogue that constitutes a particular LLM lesson in hand. This dialogue is about 1,200 words in length and has been created with a view to its interest and good humor. The dialogue of each lesson is continuous in terms of story-line with the dialogues of all the other lessons.

The dialogue is printed on the left hand half of each textbook page, a partial sentence segment, usually a sense group, per line. Partial sentence segments are of from 2 to 10 syllables. This phrasing is designed to match the rhythm of the music that will accompany the teacher's reading of the dialogue aloud. In the active concert session the students also have a line-by-line translation of the text before them.

The students are seated in the comfortable relaxizer chairs with which LLM classrooms are furnished. The teacher starts the tape recorder. The music is baroque, lively and melodious. For a moment or two, everyone sits relaxed and listens to the music. Once everyone is into the mood of the piece, the teacher begins to read the text of the dialogue aloud in keeping with the rhythm of the music. The intonation is solemn and the diction is clear. The music and spoken dialogue are of equal volume. As the teacher reads and the music plays the students follow the dialogue with their eyes, and when they don't understand they are free to glance at the translation.

The pseudo-passive concert session always follows the active concert session after a very brief break. It differs from the active concert session in two ways. In the pseudo-passive concert session the students have neither the printed dialogue text nor its translation in hand, as they had during the active concert session. Also, in the pseudo-passive concert session, the music that accompanies the teacher's reading of the dialogue aloud, although still baroque, is more austere and reflective in counterpoint and tempo than was the music used in the active concert session. The students relax in their special chairs and enjoy the music as the teacher reads the text to them. They are free to pay attention mainly to the music or to the dialogue, as they may wish. The only instruction they are given is the injunction not to let themselves fall asleep.

During the pseudo-passive concert session it is intended that subliminal learning, or learning that is peripheral to conscious attention, will take place. According to modern research, subliminal learning exhibits a very different retention pattern than does learning that has occurred during conscious-centered

attention. Conscious-centered learning is forgotten more and more with the passage of time according to the classic Ebbinghaus curve of forgetting. Subliminal learning, on the other hand, is remembered better and better with the passage of time, or so modern research suggests.

The following is an actual student's testimonial to his experience in a 5-week Spanish course:

> "The subjective reaction I had to my experience of the concert stage in the LLM course I took was not initially very encouraging. On my way home after my lesson on the days that it ended with the two concert sessions, I very frankly couldn't honestly say that I could remember much of anything about the dialogue that had accompanied the concert. Not only had I not experienced super-learning, I hadn't experienced any learning at all— or so I thought. It wasn't till my involvement in the next and final stage of each lesson, the elaboration stage, that I realized that I had learned something during the concert stage, and much more than I thought I had learned. During the elaboration stage I found to my surprise that words that I needed often came to mind. When they didn't, I'd substitute English words, which always triggered my teacher to suggest the Spanish words I needed. When he supplied the words for which I had been groping, they seemed to be old friends and easy to remember. I am sure that I remembered them much more easily than I would have been able to do had I not experienced the concert stage."

Stage 3: Elaboration

The elaboration stage is in two increments, the first coming during the lesson the next day after the pseudo-passive concert session and the second coming the next day after that. The first increment is called the primary elaboration and the second is called the secondary elaboration.

A) Primary Elaboration

In the primary elaboration students volunteer to take various roles in the dialogue. Without the help of the translation used during the active concert session, each reads his role aloud and then translates it. When a student makes a phonetic mistake while reading the foreign language text, the teacher doesn't interrupt him. The teacher lets him finish and then rereads the passage aloud himself. This is in keeping with the LLM way of handling corrections. When a student gets stuck in his translating, the teacher may suggest a word or phrase to get him started again. If the student makes an error, the teacher lets him finish and then translates the passage himself correctly. He never stops a student and tells him he's wrong.

Once the students have read and translated the entire dialogue, or as much of it as the teacher wants, the teacher brings out a light plastic ball, about the size of a basketball, for a game of catch. This is essentially a mind-focusing exercise.

To play catch, the students all remain seated in their special chairs, which are arranged in a semi-circle facing the teacher, who is also seated in a relaxizer chair. As played in primary elaboration, the teacher asks a question based on

the dialogue, observes the body language of his students, picks one who seems ready to participate, and throws him the ball. Throwing someone the ball is the LLM teacher's way of calling on someone. The student answers the question and throws the ball back to the teacher. If the student got the answer wrong, the teacher will ask the question again and throw the ball to another student, until he eventually gets the right answer. When the teacher gets the right answer, he often repeats it -- so everyone knows it was a right answer. Sometimes the teacher sets the game up so that the students toss the ball back and forth to each other. When a teacher catches the ball, he answers the question the thrower asked, asks a question himself, and throws it to another student.

B) Secondary Elaboration

The secondary elaboration of a particular lesson involves a much greater variety of activities than does its primary elaboration. The secondary elaboration of a particular lesson follows the primary elaboration by at least a day and may last for several days. Secondary elaboration activities include role playing, singing, various games, drawing, sculpting, even mild calisthenics. In all these activities language learning is present, but it is present as a component part of a greater whole and not as the whole activity upon which attention is focused.

To illustrate how secondary elaboration works, perhaps the following student's report of his experience in an LLM Spanish course may help:

"The teacher used to have us volunteer to take parts in a skit we would invent as we went along: 'Who would like to be a dentist and who would like to have a toothache?' Then he'd pass out a play telephone to the volunteers, and the 'patient' would call the 'dentist' for an appointment. The rest of the class would look on as the participants invented their conversation. Sometimes, in these skits, our teacher would pass out different hats to match the roles we chose to play. The 'taxi driver' would have one hat and the 'businessman' who wants transportation to his hotel would have another. "Once our teacher distributed all kinds of used clothing to the members of the class and has us take turns playing the role of a beggar. The person whose turn it was to be a beggar would go from classmate to classmate along the semi-circle of relaxizer chairs and beg for the clothes.

"Once, when we were learning the future tense, our teacher had us imagine that the plastic ball was a crystal ball. We took turns playing the role of fortune teller and would predict what was going to happen.

"Sometimes our teacher had us listen to a song on a tape, gave us a hand-out with the words on it, had us stand in a circle, sing it with him, and imitate his dramatic gestures.

"To learn the names of the parts of the body, we played a game. To start off, each of us chose a part of the body to be responsible for. The teacher would then call out the word for a body part, for example, 'forehead.' We all would then snap our fingers twice, clap twice, and touch our forehead twice. The person responsible for 'forehead' would then call out the word for another body part, and the process repeated. "Another game our teacher used to teach us the parts of the body was 'Simple Simon Says.' I would take time

to describe it as I imagine you all remember it from your childhood. "Even learning verb conjugations under LLM was fun and games. Our teacher had the walls of our classroom plastered with posters. Each advertised a verb conjugation or other paradigm. Each day he'd add a couple of new posters, and, if space were needed to make room, he'd remove some old ones. These products of magic marker art added lots of color to the classroom. Besides the linguistic messages, the posters included drawings of landscape scenes, people dancing, and other attention getters. To learn a conjugation we'd toss the plastic ball back and forth to each other about the semi-circle. When the ball came to you, you caught it, called out the next verb form, and tossed the ball on to someone else. If you didn't know the right form, the poster was there to look at. No one ever felt 'on the hook.'"

As defined by Dr. Lozanov himself, Suggestology learning is the comprehensive science of suggestion in all its aspects (psychological, physiological, sociopsychological, psychotherapeutical, psychohygienic, pedagogical, artistic, cybernetical, and genetic). Neverless, for the time being Suggestological learning deals mainly with the possibliities of suggestion to tap man's reserve capacities in the sphere of both mind and body. Consequently it is the science of the accelerated harmonious development and self-control of man and his manifold talents.

Suggestological learning has developed as an attempt to translate the ancient and perennial serching to tap reserve capacities, genetically predetermined in man, into a modern reality. It combines desuggestive-suggestive communicative psychotherapy with the liberating and stimulating aspects of art and some modifications of the old schools of concentrative psychorelaxation. The experimental research and the new theoretical meaning given to the phenomena researched have led to definite psychophysiological conceptions. In close connection with this research, the following three inseparable psychophysiological fundamantal principles of Suggestology have been formulated: (1) Interpersonal communication and mental activity are always conscious and paraconscious at the same time; (2) Every stimulus is associated, coded, symbolized and generalized; (3) Every perception is complex.

Applications

Presently the Lozanov Learning Method is being applied to several critical areas of educational need:

- Bilingual education in the public school system
- Basic skill training (reading, mathematics, survival skills) in the public school system
- Gifted and talented children program
- Special application to Esperanto and other linguistic areas
- Refugee training in English as a second language and cross-cultural training
- Educational development for emotionally disturbed children
- Educational and training program for children of migrant and transient workers.
- Ease of learning of rote material by adults such as medical and law students, computer programmers, etc.

Presently the LLM is in use at the Defense Language Institute, Monterey, California; U.S. Army Special Forces, Ft. Devens, Massachusetts; and at other military and governmental agencies. The results of these applications, primarily in the field of foreign languages have in most cases validated the claims

made for this method of rote learning. One of the key benefits that was also noted was the savings of time, money, and personnel. For example if a Russian or Chinese language course could be taught in 90 days as contrasted with the current course time at the Defense Language Institute of 12-18 months, considerable savings would result and also the student participants would be available for assignment for at least an extra year.

Another special application within the military would be applications in training in the complex weaponry and technical manuals with which most every soldier or sailor is familiar. A good case in point is the large amount of rote material that must be learned by cadets and midshipmen at the U.S. service academies. As a former midshipman at the Naval Academy, I can personally attest to the large amount of rote material to be learned but I also noted that the "forgetting rate" for such material was inordinately high. Presently the service academies, especially West Point are being criticized in "stretching the cadets thin."[1] I am not sure if this means their minds or bodies. In the case of the mind we are actually only using less than 10% of our brain and through this learning method we hope to tap in to that unused 90% of our brain capacity. By doing so, the modern soldier can hope to keep pace with the increasing complexity of future warfare.

It may be appropriate to conclude with the statement made by the philosopher George Leonard that "The ultimate creative capacity of the brain may be, for all practical purposes, infinite."

[1]West Point Makes a Comeback, Time Magazine, November 4, 1985, pp 28-32

Determinants of Task Effectiveness in Work Groups

Gregory P. Shea
Wharton
University of Pennsylvania

Richard A. Guzzo
Psychology Department
New York University

Abstract

Review of relevant literature has led to a new theoretical framework for understanding, researching, and managing the effectiveness of groups: managerial and non-managerial, temporary and permanent. Three variables are presented as key determinants of group effectiveness: task interdependence, outcome interdependence, and potency. Implications for theory, research, and practice are drawn.

Introduction

Groups profoundly affect organizational performance. Peters and Waterman (1982) stated that "small groups are, quite simply, the basic organizational building blocks of excellent companies" (p. 126). Groups are key factors in organizational performance generally (e.g., team CEO's, project teams, task forces, and quality circles) and in the air force specifically (e.g., flight crews, support groups, space crews and maintenance crews). Nevertheless, relatively little research has been done on group task effectiveness in organizations (cf., Gladstein, 1984; Hackman & Morris, 1975). This paper presents a theory of work group task effectiveness, one currently under test.

A Theory of Group Task Effectiveness

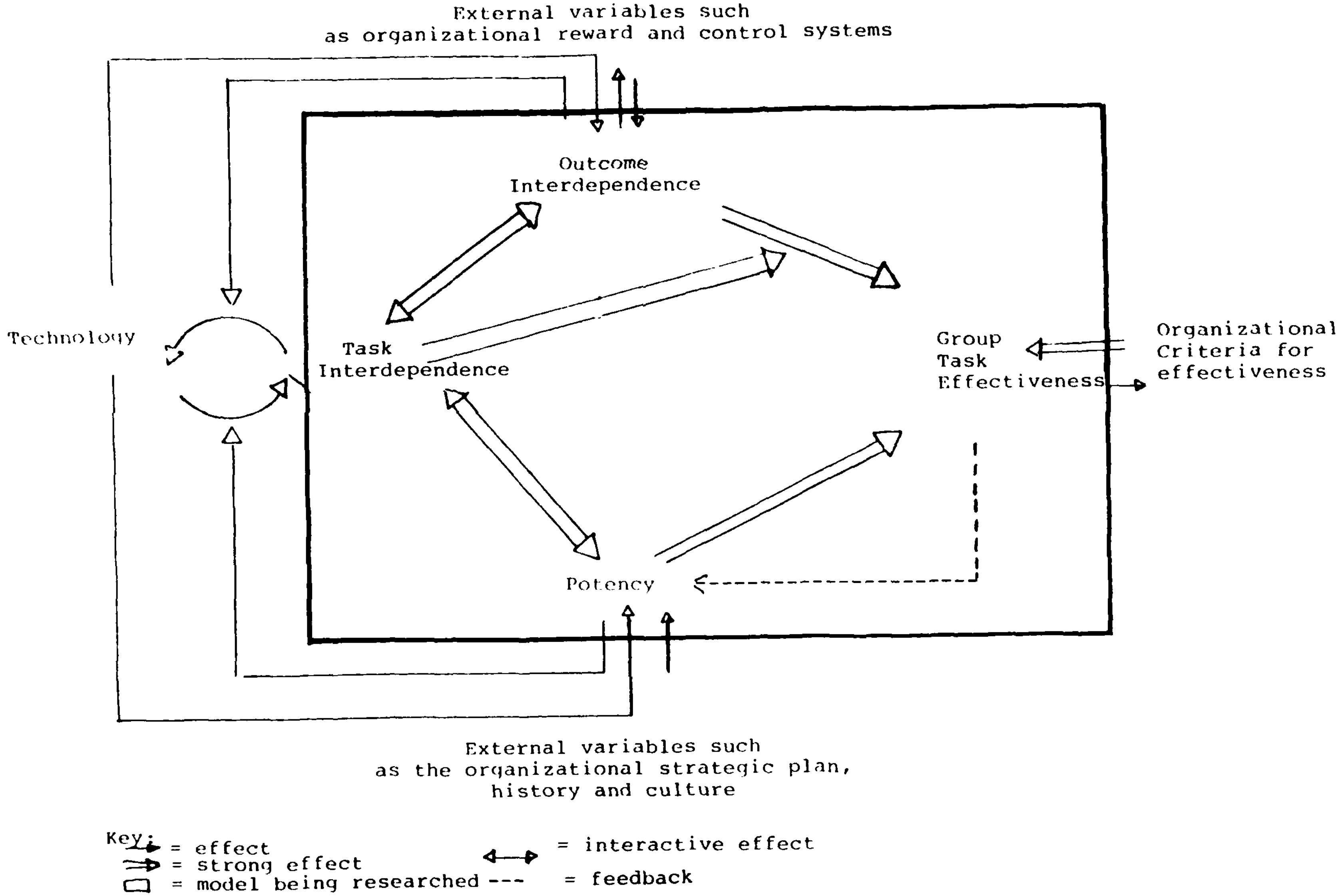
Here, the focus is on formal groups in organizations and effectiveness refers to the delivery of designated products and services per specification. Excluded from the definition are the socio-emotional consequences of group action of concern to other theorists (e.g., Cummings, 1978; Hackman, 1982). A theory proposed by Shea and Guzzo (1985) emphasized three variables as stong determinants of work group effectiveness: (1) task interdependence, the degree of task-driven interaction among members; (2)outcome interdependence, the degree and type of distribution of shared consequences of group performance; and (3) potency, the collective belief of a group that it can be effective. Figure 1 summarizes the theorized relationships among these three variables and effectiveness.

Task Interdependence

Review of the theoretical perspectives of socio-technical, job design, organizational effectiveness, and the social psychology of groups led Shea and Guzzo (1985) proposed that task interdependence

Figure 1

A Model of Determinants of Work Group Effectiveness



influences work group effectiveness in two ways. First, it moderates the impact on task effectiveness of outcomes earned by the group. Second, it affects a group's sense of potency which in turn affects group performance. Task interdependence is viewed as malleable, changing in response to factors like technology, changes in group membership, group norms, role negotiations, and decisions about how the group carries out its work.

Outcome Interdependence

Outcome interdependence exists when task accomplishment by a group yields consequences (e.g., money, survival, or the esteem of others) that are shared by group members. The present theory concerns only those outcomes (rewards or punishments) contingent on task accomplishment by a group, not outcomes predicated solely on social interaction or on the mere occupancy of a position in an organization. The outcomes of interest are those originating outside the group (such as from management or other groups).

Outcome interdependence has two important aspects: degree and type. The degree of outcome interdependence in a work group depends on the importance of outcomes contingent on the successful completion of the groups task. The greater the importance, the higher the degree of outcome interdependence. The theory presented by Shea an Guzzo (1985) assumes that a minimal degree of outcome interdependence is necessary for groups to sustain effective performance. Further, the theory posits that two dimensions characterize the type of outcome interdependence. The dimensions are competitiveness and equality in outcome distribution. These dimensions, while conceptually distinct, tend to be related in practice.

Research shows that the degree of task interdependence moderates the effect of type of distribution of group outcomes and thereby group performance. Miller and Hamblin (1963) addressed this issue when they spoke of producing versus blocking behavior within a group. Producing behaviors are those which directly contribute to the accomplishment of a group task. When outcomes contingent on group performance are distributed to group members noncompetitively or equally, each member has the same incentive to produce. When such outcomes are distributed competitively, each member has an incentive not only to produce but also to exceed the level of fellow workers in order to obtain the larger share of the competitively ddistributed reward. Exceeding the performance of others can be achieved through hindering others by engaging in through hindering others by engaging in blocking behaviors (e.g., withholding resources, intercepting clients, failing to help, or sabotage), behaviors that undermine overall group productivity. The opportunity for such behaviors is greater when task interdependence is high than when it is low. Thus, competitive distributions of group outcomes when task interdependence is high can inhibit the productivity of a group. Equal, noncompetitive distributions of group outcomes in such circumstances, in contrast, has no such effect. Shea and Guzzo (1985) discussed the existence of a third type of behavior that influences group effectiveness: facilitating. Facilitating behaviors are those which one member

performs to assist another member's producing behavior. The opportunity for and importance of facilitative behavior increases as task interdependence increases.

Finally, the type of distribution of group outcomes affects the level of task interdependence. Over time, noncompetitive distributions of outcomes are likely to engender higher task interdependence than are competitive distributions. Conversely, competing for rewards is likely to induce working arrangements that call for decreased interaction among group members.

Potency

The third primary causal factor in the proposed theory is potency, the collective belief of a group that it can be effective. This concept is similar to that of self-efficacy which Bandura (1982) defined as "self appraisals of operative capabilities" (pp. 122-123) that affect the energy, persistence, and direction of individual behavior. Bandura (1982) spoke of individual behavior and the potential usefulness of the efficacy construct at the collective level. Sayles (1958) touched on the notion of potency in groups and found it related to patterns of group behavior. The theory presented by Shea and Guzzo (1985) posits that a) potency is an important determinant of work group effectiveness, b) it is a function of, among other things, perceptions of available resources, management support and member skills, and c) groups need a minimal level of potency in order to work effectively.

The theory also predicts a mutual influence between potency and task interdependence. The greater the task interaction among group members, the greater the members' familiarity. That is, if the amount of task interaction is high, members are likely to know well each other's skill, perceptions of organizational support, and other matters that shape the sense of potency. Task interdependence thereby provides a medium for influencing potency and consequently changes in task interdependence are likely to coincide with changes in potency. Likewise, potency is predicted to influence task interdependence, since members' beliefs about each other's competence are likely to affect how members arrange themselves to do the work of the group.

Shea and Guzzo (1985) also proposed that effectiveness influenced a group's sense of potency. This proposition draws on work showing that performance feedback can powerfully shape expectations about future success on a task (Nadler, 1979).

Data and Discussion

Elements of the theory are being tested in a variety of field settings. Preliminary findings from 83 groups in the first setting generally support the theory presented above, especially the relationship between effectiveness and both potency and outcome interdependence. Also, these data generally support hypothesized relationships among the theorized determinants of effectiveness.

The theoretical framework can serve as a diagnostic device to assess the state of existing work groups in organizations and reasons for their attained effectiveness. Measures of the critical variables also can provide feedback to work groups that may stimulate them or their supervisors to change specific features of the group or its context to increase effectiveness. Further, the framework proposed here may serve as a useful guide in the design of new work groups. Research within the present framework should, consequently, attend both to theoretical and practical issues.

References

Bandura, A. (1982). Self-efficacy mechannism in human agency. American Psychologist, 37, 122-147.

Cummings, T. G. (1978). Self-regulating work groups: A socio-technical synthesis. Academy of Management Review, 3, 625-634.

Gladstein, Deborah. (1984). Groups in context: A model of task group effectiveness. Administrative Science Quarterly, 29, 3, 499-517.

Hackman, J. Richard (1982). A set of methods for research on work teams. Technical Report #1, Reserach Program on Group Effectiveness, Yale School of Organization and Management, December.

Hackman, J. R., and Morris, C. G. (1975). Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. In L. Berkowitz, (ed.), Advances in Experimental Social Psychology (Vol. 8). New York: Academic Press.

Miller, L. K., and Hamblin, R. L. (1963). Interdependence, differential rewarding and productivity. American Sociological Review, 28, 768-778.

Nadler, D. A. (1979). The effects of feedback on task group behavior: A review of the experimental research. Organizational and human performance, 23, 309-338.

Peters, T. J., and Waterman, R. H. Jr. (1982). In search of excellence: Lessons from America's best-run companies. New York: Harper and Row.

Sayles, L. R. (1958). Behavior of industrial work groups. New York: John Wiley and Sons, Inc.

Shea, Gregory P. and Guzzo, Richard A. (1985). Work group effectiveness: A theoretical view. Working Paper #11, Fishman-Davidson Center, Wharton.

"Excellence in Tactical Fighter Squadrons"

Captain Hugh A. Forde, USAF

Naval Postgraduate School*

ABSTRACT

The purpose of this research was to identify the attributes of excellence in the top performing tactical fighter squadrons.

This research followed a two-part qualitative methodology. Part I consisted of interviews with 51 general officers and colonels assigned to key leadership positions in tactical air command, e.g., CINCTAC, Vice CINCTAC and wing commanders. Each officer was asked to specify his vision of an excellent tactical fighter squadron, and to identify squadrons he considered excellent. Findings indicate that TAC's top leaders share a common vision of excellence which goes beyond monthly statistics and encompasses a broad range of qualitative issues. Four squadrons identified as excellent by consensus of the senior officers were visited and a cross section of personnel interviewed. Findings indicate that the excellent squadrons do have identifiable similarities in their method of doing business. Further research will be necessary before one can infer cause and effect based on this study, however, these findings serve as a starting point for those interested in what the excellent squadrons are doing. Further, it provides a candid, informal look into the thinking of TAC's senior leaders.

INTRODUCTION

In a recent bestseller, In Search of Excellence: Lessons From America's Best Run Companies, the authors, Thomas J. Peters and Robert H. Waterman found that the top performing American companies where characterized by certain common attributes of excellence. Does the same hold true for military units, in particular, Air Force tactical fighter squadrons? In other words, do the top performing squadrons possess common identifiable attributes which set them apart from the average squadron? If so, what are those attributes?

Unlike business corporations where success can be judged in terms of growth and the bottom line dollar figures on financial statements, tactical fighter squadrons can only be truly judged based on performance in combat. Consequently, in today's peacetime environment, one has to look at other indicators to discriminate the top performers.

Which squadrons are excellent? Who should be the judge of excellence and what criteria should be used? I chose to let a consensus of TAC's most senior, experienced officers answer these questions since, short of combat, it is the views of these senior leaders and policymakers which determine excellence. How is excellence achieved? The best qualified to answer this question are those squadrons deemed excellent by their senior leaders.

---

*This is part of a broader study of excellence in military units (Army, Navy, AF and Marine) headed by Dr Reuben Harris at the Naval Postgraduate School. All funds were provided by Naval Military Personnel Command (NMPC-6)

METHODOLOGY

Fifty three senior TAC officers were contacted for interviews, based on the following criteria: grade of O-5 or higher; prior experience as a fighter squadron commander and/or currently serving in a fighter squadron evaluator position, e.g., Inspector General Staff, wing director of operations; and currently assigned to wing level or higher, i.e., Air Division, Numbered AF headquarters or HQ TAC. One hundred percent of the officers contacted gave permission for interviews, although two were unavailable due to TDY. The interview sample included 9 general officers (O-7 to O-10), 32 colonels (O-6), 10 of whom were tactical fighter wing commanders, and 10 lieutenant colonels (O-5).

Interviews were unstructured, lasting from 30 minutes to 2-1/2 hours, centering on the question of what each officer looked for in judging a squadron to be excellent. Interview comments recorded on tape were later content analyzed to determine the main issues stressed by each officer. These were combined and summarized as "The Excellence Criteria".

Each officer was also asked to identify excellent squadrons. Although each officer had a clear vision of excellence, the majority (35 officers) gave only one or two nominations. Ten officers nominated three or four squadrons while six officers declined to identify any excellent squadrons as they felt they were too far removed from day-to-day squadron activity. However, based on the nominations given, I was able to identify seven excellent squadrons, four of which were selected for visits based on time and TDY funding limitations, and a desire to visit squadrons in different locations, flying different aircraft.

Each unit visit lasted about 2 days. A total of 42 persons were interviewed, representing a cross section of squadron personnel. The commanders' interviews lasted 1 to 2 hours while all others averaged 30 to 50 minutes. Interviews were unstructured, centering around the question of what factors are responsible for that squadron being considered excellent.

RESULTS

Part I--The Excellence Criteria--Interviews With TAC's Senior Leaders

TAC's leadership is unified, clear and enthusiastic in their vision of an excellent tactical fighter squadron. They were in total agreement that the level of performance across TAC is extremely high in all squadrons. Nevertheless, they also agreed that there are some fighter squadrons which are clearly outperforming the others. The senior officers did not define excellence in terms of any one single item, instead, each gave several criteria used in judging excellence.

Not surprisingly, squadron leadership emerged as the single most important factor they look for in evaluating whether a squadron is excellent. Daily mission performance was another key factor. But it's more than just grinding out the monthly statistics.

The excellent squadron is able to produce superior statistics under difficult conditions, but all the while maintaining consistently high standards, and a strong espirt de corps and morale; they make it look easy. The senior officers call this executing taskings with "A Touch of Class."

But what about other agents which one might expect to be factors in achieving excellence? For example, does luck or squadron location play a part? What about the type of aircraft flown and the squadron's mission? In other words, does an air-to air unit flying the newer high tech single seat aircraft have an edge over other squadrons?

The answers to these questions from the 51 senior officers interviewed was a consistent and definite NO! Each acknowledged that luck, location, type of aircraft and mission flown could be barriers to excellence but only if the squadron leadership allowed it to be so.

In a nutshell, the excellence criteria used by the senior officers in judging excellence can be summarized as follows:

Squadron Leadership: This starts with a commander who is dedicated to the squadron, actively involved and willing to make the tough decisions. Adjectives like "Out Front", "Enthusiastic" and "Visible" were contrasted with "Storekeeper", "Ghost" and "Absentee Landlord" in describing leadership in the excellent squadrons. The commander is supported by a cadre of quality intermediate leaders.

Mission: Professional Execution: The squadron has a sense of mission, challenging goals are established, training is realistic, priorities are set and there is attention to detail. The result is a high level of mission performance. But more than good statistics is the fact that the excellent squadron can forge these results and still have fun.

Mutual Support: The attitude is, "We are only as good as our weakest link." So everyone cooperates in supporting each other and the squadron, working as a team. This includes maintenance. Where there is an excellent squadron, the senior officers expect to also find an excellent maintenance unit, the two working hand in hand.

Thunderbird Syndrome: Like the Air Force thunderbirds demonstration team, the excellent squadron looks sharp! Facilities housekeeping and personal appearance reflects pride, a pride that goes beyond the individual - it's pride in the squadron. But, can a shiny exterior be a whitewash of poor performance in the air? No, looking good is not an end in itself, it is a result of being good.

Part II--The Facets of Excellence--The Story of Four Excellent Squadrons

In terms of the basic raw material, i.e. people, aircraft and facilities, the excellent squadrons are not significantly different from other squadrons. The differences are in how these resources are utilized. There is no magic or complicated formula in the way the excellent squadrons operate.

Perhaps the greatest lesson to be learned can be stated as follows: challenge people with high standards of performance, instill in them a sense of pride, show them what you want and then step back to allow them the flexibility to be creative. Above all, show them you care.

What I found in the excellent squadron was strikingly similar to what the senior officers said I would find. The words were slightly different but the basic ideas were the same:

Excellent Squadron Equals Excellent Commander. The commander is the key. A leader in the air as well as on the ground. He has a plan and knows where the squadron is going. Leadership is by example and he is respected by his people. The focus is on making the squadron the best it can be.

A Winning Attitude - They talk like winners, they act like winners and indeed they are winners. The winning attitude is an expectation of excellence which is instilled in each member from day one. The winning attitude is a concern with being the best at all times.

Combat Capability is First Priority. The training is safe but aggressive, with an emphasis on realism. New, creative ideas on better training are encouraged and given a chance to flourish. Combat capability is the why behind everything.

Teamwork and Caring; The "We" Attitude: Group effort is the way of business whether socializing or training. The squadron team is more than the aircrews, it includes enlisted as well as maintenance personnel, all working mutually rather than independently toward common goals.

Leadership Everywhere: Strong leaders permeate all levels of the organization. Qualified individuals are delegated the authority and given the flexibility to make decisions. Leaders are selected based on proficiency and demonstrated ability.

Heritage and High Standards: Excellence is a matter of tradition. The memory of squadron heroes is kept alive. Others have left a heritage of excellence which is emphasized and continued by present squadron members.

SUMMARY/RECOMMENDATION

Excellence is not just being good in one or two areas. Excellence denotes high standards across the board. As one wing commander defined it, the excellent squadron is "... like a diamond, a crystal or some fine thing, you can turn it around and look at it from many different directions, and it always looks beautiful!"

Although all the facets of excellence are important, the extensive emphasis on heritage and traditions in the four excellent squadrons visited, may be the most significant finding. Heritage and tradition enhances unit pride, gives a sense of distinctiveness and builds cohesion. It remains a potent reminder of the high standards of the past, instilling in each member a sense of history and his/her responsibility to continue that tradition of excellence. But heritage and traditions must be passed on.

I recommend that squadron commanders consider the perpetuation and maintenance of squadron heritage as a most important responsibility.

REFERENCES

1. Canan, James W. "Coming On and Coming Up", Air Force, v. 67, January 85

2. Chenette, Richard D., and Gullickson, Gregg G. Excellence In The Surface Navy, Masters Thesis, Naval Postgraduate School, Monterey, CA 1984.

3. Collins, Eliza G.C. Executive Success: Making It In Management, John Wiley and Sons, Inc., 1983.

4. Frandsen, Herbert L., Hoopengardner, David A., and Simonsen, Jerry A. Excellence In The Combat Arms, Masters Thesis, Naval Postgraduate School, Monterey, California, 1984.

5. Goldman, Nancy L., and Segal, David R. The Social Psychology of Military Service, Sage, 1976.

6. Johns, John H., Dr. and others. Cohesion In The U.S. Military, National Defense University Press, 1984.

7. Kellett, Anthony. Combat Motivation: The Behavior of Soldiers In Battle, Kluwer-Nijhoff Publishing, 1982.

8. Peters, Thomas J. and Waterman Jr., Robert H. In Search Of Excellence: Lessons From America's Best-Run Companies, Harper and Row, 1982.

9. Puryear Jr., Edgar F. Nineteen Stars, Coiner Publications, Ltd., 1971.

10. Richardson, F.M. Fighting Spirit, Crane, Russak and Company, Inc., 1978.

Enhancing Small Group Cohesion and Effectiveness in Long Range Reconnaissance Teams

Robert J. Pleban, Patrick J. Valentine
and
Thomas J. Thompson
U. S. Army Research Institute for the
Behavioral and Social Sciences
Fort Benning, Georgia

Abstract

Research was conducted to evaluate whether small group cohesion and team mission effectiveness could be enhanced by creating teams which are initially compatible on relevant interpersonal qualities. Four 6-man Special Forces reconnaissance teams were observed over a 10-day reconnaissance field exercise. After the exercise, the teams were administered the Fundamental Interpersonal Relations Orientations-Behavior (FIRO-B) questionnaire to assess team compatibility in specific need areas. Measures of team cohesion and performance were also obtained. While the evidence was not conclusive, the pattern of results suggest that the construct of compatibility, as assessed by the FIRO-B, may be an important mediator of both team cohesion and performance. Moreover, the FIRO-B may offer a cost effective means of assembling small groups with the capacity for rapidly developing into highly cohesive and effective units.

Introduction

The stresses imposed by combat and other exotic environments (e.g., polar, space, underwater) on small groups (e.g., 3-6 individuals) requires that group members function as a cohesive unit if they are to successfully complete their mission. While effective strategies exist for molding cohesive units (see Henderson, 1985), they generally take time to implement. It may, however, be possible to speed the cohesion building process by carefully selecting and matching individuals on the basis of certain interpersonal qualities, individual needs, etc. and assigning those individuals who "match up" to the same unit. The underlying assumption being that groups who are initially compatible on relevant qualities will develop more readily into a cohesive and effective unit.

A preliminary review of existing measures of personality and general interpersonal styles produced three candidate instruments for initial evaluation. These instruments included the California Psychological Inventory or CPI (Gough, 1968); Myers-Briggs Type Indicator or MBTI (Myers, 1962); and the Fundamental Interpersonal Relations Orientations-Behavior or FIRO-B (Shutz, 1958). The objective of the present research was to examine the potential utility of the FIRO-B as a means of enhancing the development of cohesion and team mission effectiveness in U.S. Army long range reconnaissance teams.

Reconnaissance teams, cohesion, and performance

Long range reconnaissance teams are frequently forced to spend long periods of time deep within enemy territory while remaining undetected. While certain

aspects of the mission can be boring or tedious, there is, nevertheless, the constant element of stress involved as the team tries to carry out its mission unnoticed, with team members often in close physical proximity for extended periods of time. Clearly, the team members must have compatible interpersonal styles if they are to work together effectively under these conditions. For these reasons, reconnaissance teams were selected as the target population for further study.

Rationale and background for the use of the FIRO-B

The FIRO-B was selected since it focuses on behaviors believed to significantly impact on compatibility, which Shutz (1958) views as being a critical mediator of small group cohesion and productivity. A summary of the major aspects of the FIRO-B is provided below.

The basic premise of the FIRO-B is that every individual has three interpersonal needs: inclusion (the need to establish and maintain a satisfactory relation with people with respect to interaction and association), control (the need to establish and maintain a satisfactory relation with people with respect to control and power), and affection (the need to establish and maintain a satisfactory relation with others with respect to love and affection). Together, these needs constitute a sufficient set of areas of interpersonal behavior for the prediction and explanation of interpersonal phenomena (Shutz, 1958).

According to Shutz (1958), inclusion is always concerned with whether or not a relation exists. Within existent relations, control is the area concerned with who gives orders and makes decisions for whom, while affection is concerned with how emotionally close or distant the relationship becomes.

The central concept used in the theoretical explanation of the interaction between individuals is compatibility. Shutz (1958) views compatibility as "... a property of a relation between two or more persons, between an individual and a role, or between an individual and a task situation, that leads to mutual satisfaction of interpersonal needs and harmonious coexistence" (p. 106).

Shutz (1958) breaks down compatibility into three major types: reciprocal, originator, and interchange. Each type of compatibility is based on the individual's expressed behavior (E) and behavior wanted (W) from others for inclusion, control, and affection as indexed by the FIRO-B scale. The scaled responses obtained from the FIRO-B are then entered into the formulas provided by Shutz (1958), yielding separate measures of reciprocal, originator, or interchange compatibilities for a set of individuals for each need area (inclusion, control and affection). These measures are described briefly in the following section below.

Reciprocal compatibility is defined as the degree to which members of a dyad reciprocally satisfy each other's behavioral preferences (i.e., does j express the behavior wanted by i, and does j respond favorably to the type of behavior i characteristically expresses). Originator compatibility is defined as the extent to which two individuals complement each other based on their tendencies to originate or initiate behavior (e.g., do those who wish to dominate and control the activities of others work with those who want to be

controlled). Interchange compatibility is defined as the extent to which individuals prefer similar amounts of exchange for a specific commodity (interaction, power, love). Both reciprocal and originator compatibility are viewed as being primarily applicable for dyads whereas interchange compatibility is more meaningfully applied to groups (Shutz, 1958).

Method

Subjects

Subjects were 24 white male soldiers (four 6 - man Special Forces teams) enrolled in the nine week Advanced Land Reconnaissance Course (ALRC) at Fort Bragg, North Carolina, who were montiored over a 10 day field exercise during the period 27 July - 7 August 1985.

Materials

Four instruments were employed. They included: 1) Fundamental Interpersonal Relations Orientations - Behavior (FIRO-B), Biographical Questionnaire (BQ), 3) Evaluator Debriefing Questionnaire (EDBQ), and 4) Subject Debriefing Questionnaire (SDBQ).

The FIRO-B scale is a 54 item inventory consisting of statements designed to tap the subject's expressed and wanted behaviors for inclusion (e.g., I try to be with people), control (e.g., I let other people decide what to do), and affection (e.g., I try to have close relationships with people). Each item is rated on a 6-point scale (usually, often, sometimes, occassionally, rarely, never) and is then combined into one of six subscales consisting of nine items (expressed inclusion, wanted inclusion, expressed control, wanted control, expressed affection, wanted affection). The BQ is a 41 item multiple choice - short answer instrument developed to provide background information on subjects' socioeconomic status, family life, career choice, army experience, social history, general mood state, and disposition. The EDBQ is an 8 item instrument consisting of 7 point rating scales and short answer questions designed for the evaluators to assess team effectiveness, cohesion and specific shortcomings demonstrated by the team during the exercise. The SDBQ is an 11 item instrument consisting of 7 point rating scales and short answer questions designed to probe team members' perceptions of how well their team performed, level of cohesion, leadership style/effectiveness, individual skill deficiencies, and suggestions for improved training.

Procedure

Prior to the actual start of the reconnaissance training mission (i.e., field exercise), soldier subjects were given a general description of the nature of the research This was followed by the administration of the BQ.

Shortly after the soldiers completed the BQ, they were isolated as team elements in specified areas where they remained removed from further outside contact for 48 hours. This time was used by the teams to prepare for the mission. The major objective of the mission was to gather information on any movement of equipment or personnel through a designated area of observation. The secondary objective was to complete the mission without being detected while moving into a designated (field testing) area, while operating in the area, and

while moving out of the area toward a pick up point (which marked the end of the exercise).

The reconnaissance teams were monitored and evaluated by instructors who were highly skilled in reconnaissance techniques. The instructors moved in and out of the teams' observation locations undetected during all hours of the day and night, checking for any sign of the teams' presence in the area. The instructors also spent some planned contact time with the teams during the movement phases of the exercise noting any violations of noise and light discipline which might alert the target military units being observed.

The exercise officially terminated at 0500 on the 10th day. Subjects were then extracted from the exercise area to Fort Bragg for debriefing. The team leaders first briefed the principal instructor, providing detailed summaries of the events of the past 10 days. At the conclusion of the leader's briefing, the instructor or evaluator who was assigned to each team presented separate critiques of the team's overall performance and of individual member performances.

The following day the subjects were given the SDBQ and the FIRO-B while the evaluators were given the EDBQ. During this time subjects and evaluators were probed by the experimenters about specific aspects of the research.

Results

Team compatibility was operationalized using Shutz' (1958) interchange measures for inclusion, control, and affection. The computed values obtained from the FIRO-B scale for expressed and wanted inclusion, control, and affection for each subject were paired with those from each of the subject's five remaining team members using the interchange formula derived by Shutz (1958). The compatibility scores for all dyads were averaged for each team for interchange inclusion, control, and affection. Teams were then ranked from least to most compatible for each measure.

Similarly, the relevant items from the SDBQ, questions 1 "How effective was your team in accomplishing its mission objectives?" and 2 "How well did your team work together?" and the EDBQ, questions 1 "How effective was the leader in making sure that both individual and group tasks were accomplished during the exercise?" and 3 "How well did the leader and team members work together during the exercise?" were scored and averaged for each team. Teams were then ranked from least to most effective/cohesive for each item.

Rank-order correlations were computed among all measures of compatibility and the four questionnaire items. Since a significant correlation at the .05 level is 1.00 (for $N = 4$), only the correlations between SDBQ 1 and the interchange inclusion and affection measures reached statistical significance. The remaining correlations were all positive (the more compatible the group the more cohesive/effective the group was rated). Seven of the 12 computed correlations were $\geq$ .95.

One notable trend was apparent. Both interchange inclusion and affection correlated strongest with SDBQ 1 and EDBQ 1 and 3. However, when the interchange control measure is considered, the correlations with the same items are all reduced 20-60 units. This pattern is reversed with SDBQ 2.

Discussion

While not conclusive, the present results are, nevertheless, encouraging with regard to using the FIRO-B as part of an overall screening process geared to optimizing the selection and assignment of individuals to specified reconnaissance teams. Since the unit of analysis was the group, sample size was significantly reduced, which had an adverse impact on statistical power. This may account, to a large extent, for the paucity of statistically significant correlations between the three measures of compatibility and the items tapping team cohesion and effectiveness. However, the size of the correlations and the fact that all the correlations were positive are noteworthy.

From an interpretive standpoint, the pattern of correlations that were obtained may shed some light into the nature of the relationship(s) between compatibility and team cohesion and effectiveness. The interchange inclusion and affection measures correlated strongest with SDBQ 1 (subject's assessments of the team effectiveness), and EDBQ 1 (evaluator's assessments of team effectiveness), and EDBQ 3 (evaluators' assessments of how well the team worked together), while the interchange control measure correlated strongest with SDBQ 2 (subjects' assessments of how well the team worked together) and EDBQ 3.

These results suggest that perceptions of team cohesiveness (SDBQ 2, EDBQ 3) may be linked to the level of agreement existing among team members as to how much control should be exhibited over them by the team leader (interchange control compatibility).

Perceptions of team effectiveness (SDBQ 1, EDBQ 1) and evaluator perceptions of team cohesiveness (EDBQ 3), however, seem to be most sensitive to the levels of agreement existing among team members as to how much interaction (interchange inclusion compatibility) and emotional closeness (interchange affection compatibility) they are comfortable with.

These interpretations should be viewed cautiously, however, because of the small N. Future research, employing a wider diversity of cohesion and performance measures (i.e., paper and pencil and behavioral) and larger samples should help clarify the interpersonal dynamics involved in cohesive/productive vs non-cohesive/non-productive small groups, and the subsequent role of the FIRO-B in the selection of long range reconnaissance teams.

References

Gough, H. G. (1968) An interpreter's syllabus. Palo Alto, CA: Consulting Psychologists Press.

Henderson, W. D. (1985). Cohesion: The human element in combat. Washington, DC: National Defense University Press.

Myers, I. B. (1962). The Myers-Briggs type indicator. Palo Alto, CA: Consulting Psychologists Press.

Shutz, W. C. (1958). FIRO: A three-dimensional theory of interpersonal behavior. New York: Rinehart.

The Effects of Combat on Small Unit Cohesion And Bonding

Stanley H. Holgate
United States Army Medical Research Unit - Fort Bragg
Walter Reed Army Institute of Research

Abstract

Small unit cohesion and bonding was measured by two participant observers while deployed with an airborne infantry company on a field training exercise. On the day before redeployment a 25-item questionnaire was administered to 62 paratroopers half of whom had participated in Operation Urgent Fury on the island of Grenada. Factor analysis of the questionnaire data yielded two primary factors related to horizontal and vertical bonding and four secondary factors related to buddyships and buddy systems. Initial analyses of the factored scale scores failed to support the notion that participation in a combat event contributes significantly to horizontal bonding. Rank and time-in-platoon seemed to be better predictors of cohesion and bonding scores. The apparent minimal effect of combat experience was determined to be due to confounding produced by individual re-assignments during the period following Operation Urgent Fury. Further analyses revealed a significant facilitatory effect of combat on horizontal bonding.

Introduction

The concepts of cohesion and bonding as they apply to the military were derived from field observations during WWII and the Korean War (Marshall, 1947; Little, 1964). Cohesion in military units is reflected in the horizontal and vertical bonds among unit members. Bonding strengths are dependent on time and space factors as well as environmental conditions. Horizontal bonding strength seems to increase as a function of the length of time a group is together, the proximity of the members, and higher stress levels. Vertical bonding seems to be more a function of individual skill levels and the unit's role and capabilities in support of the mission. Long term vertical bonding may have as much supportive power as horizontal bonding.

Presented here is an effort to demonstrate the relative contributions of horizontal and vertical bonding to cohesion in an airborne infantry company which had deployed on a real world combat mission, Operation Urgent Fury, almost a year earlier. The unit was among the first to arrive on Grenada and was among the last combat units to leave the island eight weeks later. Half of the troops surveyed in this report had deployed to Grenada. The remaining

troops had joined the unit after it returned from Grenada. Based upon Little's research, stronger horizontal and vertical bonding scores were expected among those troops who had deployed on Operation Urgent Fury.

Method

Two participant observers, a major and a private first class, deployed with an airborne infantry company on a field training exercise (FTX). The participant observers were introduced to the company prior to the deployment and attached to selected squads. The participant observers deployed with the advance party and joined their respective squads on the drop zone. They remained with their squads through redeployment. A short questionnaire consisting of general information items and 25 cohesion/bonding items was administered to 62 paratroopers (squad members) who were present on the day before redeployment. Thirteen items had been selected from an existing unit perceptions questionnaire which is routinely used to assess unit cohesion. Twelve items had been included for the purpose of assessing the nature of buddyships and buddy systems within the unit.

Results

Factor analysis of the questionnaire data using Varimax rotation resulted in two primary factors and four secondary factors (Table 1). A cluster

Table 1. Factors derived from 25-item questionnaire.

HORIZONTAL BONDING (Variance Explained: 4.591)	FACTOR LOADING	BUDDY SUPPORT (Variance Explained: 1.997)	FACTOR LOADING
I really enjoy being a member of this squad.	0.873	Everyone needs a buddy with whom to talk about problems.	-0.678
People really look out for each other in this squad.	0.859	Who do you spend time with after duty hours?	-0.663
Almost all of the people in this squad can really be trusted.	0.799	I don't have time to be anyone's buddy.	0.626
People in this squad feel very close to each other.	0.774	If my buddy needed it, I would loan him $50, no questions asked.	-0.618
I never have just one buddy because in my squad all are buddies to one another.	0.656	**BUDDY NUMBER** (Variance Explained: 1.573)	
If we went to war tomorrow, would you feel confident going with this squad or would you rather go with another squad?	0.575	There are a lot of guys in the squad I consider to be my buddy, but there is only one with whom I discuss personal problems.	-0.712
It's hard to have a buddy in this unit because everyone goes his own way at the end of the day.	-0.554	Soldiers living in the barracks have more of a buddy system than soldiers living off post.	-0.659
VERTICAL BONDING (Variance Explained: 3.177)		**BUDDY SELECT** (Variance Explained: 1.765)	
How often, aside from meetings, does your platoon leader talk with you personally?	0.769	I usually find a buddy only when the squad goes the field.	0.560
... your platoon sergeant....	0.710	There is usually someone in my squad whom no one wants for a buddy.	0.441
... the company commander....	0.668	Married men seem to buddy only with other married men.	0.417
If I have to go into combat, I have great confidence in my personal skills and training.	0.565	**BUDDY NEED** (Variance Explained: 1.606)	
I am impressed by the quality of leadership in this platoon.	0.560	I prefer being alone most of the time, but I like having a buddy on whom I can count in time of need.	-0.845
Is your squad leader ever included in after duty activities?	0.536	If my buddy was injured in combat, I would continue with the mission.	0.491
I think the level of training in this squad is very high.	0.455		

analysis of the factored scale scores yielded two main branches on a dendrogram (Table 2). One branch contained the two primary factors, horizontal bonding (HORBOND) and vertical bonding (VERBOND). The other branch consisted of diverging limbs containing the secondary factors relating to buddyships and buddy systems. For this report detailed analysis of the data is limited to horizontal and vertical bonding as well as cohesion which is represented by BUDBOND (sum of the six scale scores).

Table 2. Cluster analysis of the factored scale scores using Euclidean distances and nearest neighbor as the linking method.

```
VERBOND  ---------------------------
                                    I-----
HORBOND  ---------------------------     I
                                         I-----
BUDNMBR  ---------------------           I
                                  I---------------
BUDSELEC ------------            I
                      I--        I
BUDSPT   ------------  I         I
                        I-------
BUDNEED  ---------------
```

Table 3. Factored scale means grouped by treatments. RANK 1 = PVT-PFC; RANK 2 = SP4; and RANK 3 = SGT. TIPLT 1 = 6 mo or less; TIPLT 2 = 7-12 mo; and TIPLT 3 = more than 12 mo. Asterisks denote which cells contain differences less than the .05 level of significance.

		FACTOR SCALE		
TREATMENT	N	BUDBOND	HORBOND	VERBOND
				*
VETS	32	52.1	14.3	12.9
NON-VETS	30	57.1	15.9	15.2
		*	*	*
RANK 1	43	57.8	16.1	15.3
RANK 2	14	49.3	13.4	12.3
RANK 3	5	40.6	10.8	7.8
				*
PLATOON 1	18	58.0	15.6	15.9
PLATOON 2	15	53.7	15.1	12.9
PLATOON 3	29	52.8	14.8	13.3
		*	*	
TIPLT 1	24	58.0	16.8	14.5
TIPLT 2	19	57.5	16.9	14.6
TIPLT 3	19	47.2	11.0	12.8

The treatment means for the VETS and NON-VETS in Table 3, suggest that the Grenada deployment had some facilitatory effect on cohesion in general and vertical bonding in particular. (Smaller scale scores reflect stronger bonding among respondents.) Participation in Operation Urgent Fury did not appear to have had a lasting effect on horizontal bonding. The expected differences between the means for the GRENADA VETS and NON-VETS on the HORBOND scale did not materialize (Table 3) whereas there were significant differences on the HORBOND scale across RANK and TIPLT. When the data were evaluated for the effect of TIPLT, significantly stronger scores were found for troops who had been in the platoon for over a year (TIPLT 3) than troops who had been in the platoon for less than a year (TIPLT 1 and TIPLT 2). Significantly stronger HORBOND scores were observed for sergeants (RANK 3) compared to specialists (RANK 2) and privates (RANK 1).

The data in Table 4 suggest that the absence of a significant difference in horizontal bonding scores between GRENADA VETS and NON-VETS resulted from confounding produced by personnel re-assignments during the period between Operation Urgent Fury and the FTX. Attention was focused on PLATOON 3 in which 8 of the 18 VETS had been in the platoon long enough to have deployed

together on Operation Urgent Fury. The remaining VETS and the 11 NON-VETS had joined PLATOON 3 after the platoon had returned from Grenada. Differences between the means of the factored scale scores in Table 5 support the notion that participation in Operation Urgent Fury contributed to the horizontal bonding strength among the personnel in the platoon who had deployed with the unit to Grenada. The data also demonstrate that GRENADA VETS and NON-VETS who had transferred into the platoon since Operation Urgent Fury had not yet developed strong horizontal bonds within the platoon. These two groups did not differ significantly on HORBOND scores, but the VETS did have significantly stronger VERBOND scores than the NON-VETS.

Table 4. Number of Grenada veterans and non-veterans by platoon and time-in-platoon.

		PLATOON			
TIPLT	GRENADA	1	2	3	TOTAL
6 MONTHS OR LESS	VETS	2	0	4	6
	NON-VETS	5	4	9	18
7 - 12 MONTHS	VETS	1	1	6	8
	NON-VETS	4	5	2	11
MORE THAN 12 MONTHS	VETS	5	5	8	18
	NON-VETS	1	0	0	1
TOTAL		18	15	29	62

Table 5. Factored scale means grouped by treatment. The data in this table represent three categories of membership in PLATOON 3. ABC 1 = GRENADA VETS who served with this platoon in Operation Urgent Fury; ABC 2 = GRENADA VETS who served in another platoon in Operation Urgent Fury but who are now members of this platoon; ABC 3 = GRENADA NON-VETS who are now members of this platoon. RANK 1 = PVT-PFC and RANK 2 = SP4-SGT. An asterisk signifies a difference among means less than the .05 level of significance.

		FACTOR SCALE		
TREATMENTS	N	BUDBOND	HORBOND	VERBOND
		*	*	*
ABC 1	8	43.4	10.3	11.3
ABC 2	10	55.7	17.5	12.3
ABC 3	11	57.0	15.6	15.7
ABC1				
RANK 1	4	44.5	9.5	12.5
RANK 2	4	42.3	11.0	10.0
ABC2				
RANK 1	5	61.6	21.2	14.8
RANK 2	5	49.8	13.8	9.8
ABC3				
RANK 1	10	58.1	16.1	15.9
RANK 2	1	46.0	11.0	14.0

The differences in VERBOND scores may be attributed as much to rank as to any influence of participation in Operation Urgent Fury. This is evident in the VERBOND means across RANK in Table 3 and the fact that the RANK by GRENADA interaction was significant. The largest difference was between RANK 2 and RANK 3. This difference is probably due to the fact that sergeants are more likely to communicate "up the chain" than are specialists. Even when SP4s' and SGTs' scores are combined in Table 5, the impact of rank on VERBOND scores for the Grenada VETS who had transferred into PLATOON 3 is easily seen.

Conclusions

An assessment of unit cohesion at the platoon level has to take into consideration both horizontal and vertical bonding strengths. Bonding strengths are dependent on rank, time in the unit, and unit participation in a significant event such as combat. The effect of even brief exposure to combat seems to have a lingering, strengthening effect on cohesion for those soldiers

remaining in the unit. But for new personnel to the unit, including veterans from other units, the combat event seems to interfere with the horizontal bonding process. Vertical bonding is not effected by combat to the same degree as horizontal bonding. Thus, it is important for commanders to know that following events such as Operation Urgent Fury a platoon's combat readiness may be degraded due to lower horizontal bonding strength among replacements.

References

Little, R. W. (1964). Buddy relation and combat performance. In M. Janowitz, (Ed.) The New Military: Changing Patterns of Organization (pp.195-223). New York: Russell Sage Foundation.

Marshall, S. L. A. (1947). Men Against Fire. New York: William Morrow.

Disclaimer

The views of the author do not purport to reflect the position of the Department of the Army or the Department of Defense, (para 4-3, AR 360-5).

Pharmacological Optimization of Combat Performance

LTC Frederick J. Manning

Division of Neuropsychiatry
Walter Reed Army Institute of Research

Abstract

This short paper provides some of the historical background and rationale for reopening the question of whether some of the stressors of the modern day battlefield may be made more manageable by judicious sampling from a burgeoning pharmacology industry

Widespread use of marihuana and heroin by U.S. servicemen in Vietnam was a practice roundly condemned by military leaders and civilians at home as an indication of bad soldiers, bad leaders, or a bad war, depending on the position of the critic. The drugs clearly served some purposes for the users themselves however. Bentel, Smith and Crim (1971) for example report on how social bonding was facilitated by communal drug use, how heroin was used by some as "therapy" for their despair and frustration, and how still other soldiers spoke of carefully titrating their use of marihuana while on combat patrols to calm down, enhance awareness, and increase their suspiciousness of enemy activity. It might be argued that these soldiers differed from their counterparts in prior wars primarily in their choice of illicit drugs for their self-medication. Alcohol in fact has a long and glorious tradition among military men as a builder of cameraderie - cohesion is the currently fashionable term. It very likely has been used for its direct effects on fighting spirit as well. The term "Dutch courage" is thought to have originated from the widespread use of Dutch manufactured gin by 18th and 19th century soldiers (Jones, 1985). When the great British general Wolseley decided to abolish the rum ration during his 1884 campaign to relieve Gordon at Khartoum, he needed the personal intervention of Queen Victoria to overcome the strenuous veto of the War Office -- and he suffered his only major set back in the subsequent campaign (Manning, 1984). The ancient Viking "berserkers," as well as some American Indian tribes, are thought to have included ingestion of psychodelic mushrooms in their preparation for battle (Aaronsan and Osmond, 1970).

The pharmaceutical industry has exploded in the last fifty years, and, though its greatest impact on the military has been in the areas of vaccines and antibiotics, "performance enhancing" drugs have not gone entirely unnoticed. The Soviet army of World War II apparently used amphetamine to stave off fatigue, as did Rommel (Cuthbertson, cited in Laties & Weiss, 1981), but other 20th century uses have been focused on limited non-shooting aspects of combat. US soldiers on "long range reconnaisance patrols" in Vietnam for example were issued methylphenidate (Ritalin) and sometimes dextroamphetamine, but with instructions to use it only for the long march back to base camp at the completion of the mission (Jones, 1985; Malone, 1984). More

recently, an Israeli physician capitalized on the sleep-inducing "side effect" of the antihistamine Dramamine to promote sleep during the long flight prior to recapturing the hostages at Entebbe airport (Dolev, personal communication, 1982). Finally, Baird, Coles, & Nicholson (1983) have reported on the widespread use of the short-acting hypnotic temazepam by British aircrews during the Falklands conflict in 1982. Flying rates were extended far beyond previous experience, particularly in reconnaisance and transport roles, and temazepam was used to insure that crews got maximum sleep in their non-flying hours.

It was with these examples in mind that the Division of Neuropsychiatry at the Walter Reed Army Institute of Research assembled a small group of scientists to examine in a systematic way the possible uses of behaviorally active drugs in military operations. Nearly all the members were sceptical that a good case could be made for any currently available drug. Some remain sceptical today, but others have slowly come to feel that under some circumstances, some drugs, at some dose level, might provide our soldiers a worthwhile edge. All agree that recent developments in neurochemistry provide substantial grounds for continued research along these lines.

The four papers included in this symposium focus on four different aspects of soldier performance: physical strength and endurance; anxiety; sleep; and cognition. These categories are not independent of each other, and certainly are not meant to encompass all behavior. Furthermore, a number of behaviorally active drugs are clearly relevent to more than one category. As a heuristic device however, this division of labor has served its purpose of organizing our considerations around required performance rather than drug structure or disease entities.

All of our Psychopharmacology Group's members were keenly aware from the start that altering CNS chemistry was something that simply cannot be done without wide ranging behavioral effects, only some of which will be desireable, no matter what the context. We thus conceded from the outset that the chances of finding or ever developing a drug that would produce a "Superman," i.e., significantly enhance the maximum performance of a well-trained, well-motivated soldier. There are however a myriad of features (stressors) of the modern battlefield which insure that nearly no one will in fact be performing at or near maximum for very long. Two of these features have been especially influential in our thinking about drugs. The sheer lethality of modern weapons for example and the anticipated need for wide dispersal and a 360° "front" lead us to expect unprecedented numbers of battle stress (i.e. psychiatric) casualties. Second, the availability of night-vision devices has made it possible for Soviet doctrine to specify <u>continuous</u> operations. Following even a rough approximation of this doctrine will quickly cause enormous difficulties for Western forces, which simply do not have the numbers to go to a simple shift work schedule. A more modest and theoretically achievable goal than building a superman has instead been to mitigate the decrement in performance known to occur in highly stressful circumstances (a la sports psychology) or as a function of increasing fatigue and sleep deprivation. There will undoubtably be "trade-offs" to be considered in using these drugs, as there are with any drug, but we see these drugs as emergency equipment, the alternative to which is death on the battlefield. It is largely for this reason also that we have not considered vitamins, steroids, ginseng, and other compounds which require chronic or long-term administration.

In summary, I do not feel I am being unfair to any of the four speakers who will follow me to say that there are no panaceas on the horizon, but there are some promising leads for repairing or for stalling deterioration of performance on the modern battlefield. We owe it to our soldiers and our country not to ignore them.

References

Aaronson, B. & Osmond, H. (1970). Psychedelics, New York: Doubleday.

Baird, J.A., Coles, P.K.L., & Nicholson, A.N. (1983). Human factors and air operations in the South Atlantic campaign: discussion paper. Journal of the Royal Society of Medicine, 76, 933-937.

Bentel, D., Smith, D., Crim, D. (1971). Drug abuse in combat: the crisis of drugs and addiction among American troops in Vietnam. Journal of Psychedelic Drugs, 4, 23-30.

Jones, F.D. (1985). Sanctioned use of drugs in combat, in Pichot, P., Berner, P., Wolf, R., & Thau, K. (Eds), Psychiatry, (Vol. 6) 489-494. New York: Plenum.

Laties, V.G., & Weiss, B. (1981). The amphetamine margin in sports. Federation Proceedings, 40, 2689-2692.

Malone, D.M. (1983). Oral presentation to Division of Neuropsychiatry, Walter Reed Army Institute of Research, Washington, D.C. (Nov 8)

Manning, F.J. (1984). Critical Commentary, in Ingraham, L.H., The Boys in the Barracks, 205-242. Philadelphia: Institute for the Study of Human Issues, Inc.

Note

The views expressed by this author and the other contributors to this panel are their own, and do not purport to reflect the position of the Department of the Army or the Department of Defense.

Pharmacological Optimization of Military Performance: Strength and Endurance

G. Jean Kant

Department of Medical Neurosciences, Division of Neuropsychiatry
Walter Reed Army Institute of Research

Abstract

A pill that could increase strength and endurance would be of considerable benefit to soldiers. Combat can be exhausting, straining physical and mental resources beyond the breaking point. Heavy loads, sleep loss, fear and anxiety, wet clothing, cold or heat, jet leg and other factors often present in a combat situation can be expected to decrease endurance and strength and further impair military performance.

Historically, various drugs have been used to attempt to increase endurance in combat. These drugs include alcohol, amphetamines, caffeine, nicotine and methylphenidate (Ritalin) in US forces and possibly other drugs in non US forces. In addition, purported performance enhancers have been used by athletes all over the world in hopes of improving strength or speed in athletic competitions. The belief that drugs can improve performance is widespread among athletes, as is shown by the need to drug test athletes (and horses) to keep athletic contests fair.

Surprisingly, given this wide acceptance of the power of drugs, very little objective evidence exists that drugs do actually improve performance. It is the purpose of this report to review various drugs from the standpoint of their usefulness in enhancing militarily relevant performance.

As a result of this review, it is concluded that currently available pharmaceuticals do not significantly improve speed or endurance in healthy rested individuals, although drugs that increase wakefulness are available. Very little testing of drugs has been conducted in fatigued subjects. Thus it is possible that some currently available drugs might be effective in preventing fatigue-induced degradations in performance or in revitalizing performance in already fatigued subjects. Drugs of the future might actually be able to improve endurance and strength and enhance performance in non-fatigued soldiers.

Introduction

The ability of the individual soldier to maintain optimal physical and mental performance over a sustained period of time is critical to the success of military operations. The urgency, the excitement, the need for action may carry a soldier through a short period on the battlefield, but inevitably the physical and mental stress of sustained combat can be expected to take its toll on performance. How can these performance decrements be prevented or slowed?

Physical fatigue can be minimized by conditioning. Training and conditioning increase the blood supply to affected muscle groups, increase muscle glycogen stores , increase muscle mitochondrial number and oxidative capacity to provide fuel for exercising muscle (1). Since fear can interact with fatigue to speed performance deterioration under stress, mental conditioning prior to stress challenge can also presumably improve performance. Adequate training reinforces coping behavior which in turn reduces stress and thus fatigue. Adequate sleep can forestall mental fatigue. When

possible, judicious use of sleep/rest cycles and jet lag countermeasures can reduce fatigue. Thus, to some extent, non-pharmacological measures can improve performance if these training techniques are implemented prior to stress challenge.

Nonetheless, drugs potentially could improve performance beyond what is achievable through training and conditioning. Alcohol and amphetamines have been issued to troops in past wars to reduce fear and decrease fatigue (5,8). The results achieved are difficult to objectively evaluate. In addition to the anecdotal nature of the data, these drugs affect mood and perception so that a soldier may feel that the drugs enhanced his performance when they actually did not.

Potential Performance Enhancers: Strength and Endurance

Amphetamines

Moderate doses of amphetamines have been reported to elevate mood, increase alertness, decrease sleep time, increase motor and verbal activity, increase anger and verbal aggression and decrease appetite (2). For various of these effects, students, truck drivers, dieters, athletes and soldiers have used and abused amphetamines since their rediscovery in the 1920's. While the euphoric, appetite suppressant and sleep reducing effects of amphetamine have been documented in scientific studies, data regarding the effects of amphetamines on endurance or speed is contradictory with various studies reporting no significant effects of amphetamines or small positive effects (6,8).

One extensive double-blind study of the effects of amphetamine sulfate on athletic performance concluded that amphetamine had no effect on athletic performance (6). This study is described in detail below because it is one of the best scientifically conducted studies available on the effects of amphetamine on human performance.

Subjects were 54 male college athletes. Each subject was tested 3 times with amphetamine sulfate (10 or 20 mg ingested orally 30 to 60 min prior to testing) and 3 times with placebo. The following activities were tested (1) running to exhaustion on a treadmill twice in succession with a 10 min rest in between (2) swimming 100 yd twice with a 10 min break (3) swimming 220 and 440 yd once on each test day (4) running 220 yd time trials (5) running distances from 100 yd to 2 miles in competition. Approximately 10 subjects were run for each of these activities. No significant differences in performance were seen in athletes on amphetamine tests versus placebo tests. Average performances were almost identical between treatment groups. Overall, no beneficial or deleterious effects of these doses of amphetamine (which were sufficient to cause some complaints of insomnia) were observed on speed or recuperation (the second swim or run). The performance of individual subjects varied with some performing "better" on amphetamine and some "worse". Some athletes attributed a good performance to the "miracle drug" that they took although later examination of the data revealed the "miracle drug" was placebo. "Most of the subjects felt a sense of euphoria after the 20 mg dose of amphetamine ... They also <u>thought that they could run longer</u> when they felt this effect, <u>but they did not.</u>"

The lack of convincing data demonstrating that amphetamine improves physical endurance coupled with the drug abuse potential of this drug makes it an unsuitable drug for field use for this purpose. If better drugs are not available, amphetamine does have potential use as a "keep awake" drug that might decrease mental fatigue for short periods. When amphetamine intake is discontinued, a rebound period of increased sleep and lack of energy can be expected.

Ritalin

Ritalin (methylphenidate) is a central stimulant with some similarity to amphetamine but with much less euphoria and abuse potential. Its main use is in the treatment of hyperkinetic children. It has been used by US troops as a stimulant (5).

Anabolic Steroids

Athletes particulary those involved in strength sports, (e.g. weightlifting) abuse anabolic steroids to the extent that drug testing for these compounds in competitions is required. Obviously, many athletes believe that these steroids build muscles and improve strength performance. The endogenous steroid, testosterone, is believed to affect the development of muscles in males at puberty. However, scientific evidence that anabolic steroids can improve strength in adults is scant. In army basic training, at a time when muscle strength and endurance increase, the levels of testosterone decrease.

The effects of an anabolic steroid (Nibal, 20 mg) on physical performance were tested in a 16 week study on 47 young men (4). Eight men received placebo only, nine men took steroid, 15 men received placebo plus exercise and 15 men received steroid and exercise. The authors of this study conclude "There were no significant differences in strength, motor performance, or physical working capacity between the steroid and placebo groups." Exercise in both control and steroid groups did increase performance on some tasks. Steroid administration also did not affect vital capacity, limb circumferance or skin-fold thickness. However, the dose of steroid used in this study was much less than the reportedly massive doses taken by some power lifters. It is quite possible that these massive doses do increase muscle mass and strength. Steroids at high doses have been shown to produce serious liver and heart damage. The potential benefit of steroids seems to be clearly outweighed by the health risk they present.

Human growth hormone (hGH)

Human growth hormone is a pituitary hormone required for normal growth in humans. Extracts of pituitaries obtained from cadavers are used to treat children with growth hormone deficiencies. Recently, some power athletes have taken hGH for the purpose of increasing muscle mass and decreasing body fat. The probable side effects of hGH administered to a person who does not have a growth hormone deficiency include acromegaly and dangerous heart enlargement. Current supplies of hGH are very scarce and insufficient for required medical uses. However, drug companies are working on synthesizing growth hormone. If supplies become available at a reasonable price, abuse of this hormone can be expected. Given the probable deleterious side effects of this drug and the long time required for an effect on muscle mass, this hormone is not militarily useful.

Caffeine

Caffeine is probably the most widely used stimulant drug. One cup of coffee contains approximately 100 mg of caffeine, a therapeutic dose. Cola beverages contain about 50 mg of caffeine. The stimulant actions of caffeine and related xanthines (theophylline, theobromine) undoubtedly contribute to the popularity of the beverages that contain them (coffee, tea, cocoa, cola).

Caffeine and related xanthines are central nervous system stimulants, reportedly facilitating rapid, clear thought, allaying drowsiness and fatigue, decreasing reaction time, increasing sensitivity to sensory stimuli and increasing motor activity (3,7).

Caffeine also has significant peripheral effects: increasing respiration, cardiac output, peripheral vasodilation, peripheral blood flow, force of skeletal muscle contraction, capacity for muscular work, basal metabolic rate, lipolysis and

glycogenolysis.

High doses of caffeine can cause insomnia, excitement, restlessness and even mild delirium. Other non-desireable effects include sensory disturbances (e.g. ringing in the ears) and muscular tremor. There is a wide range of individual sensitivity to caffeine with some individuals reacting to very low doses and others (especially chronic users) developing tolerance to caffeine.

Caffeine use is self-prescribed for the most part as people reach for another cup of coffee to keep them alert to study or drive or man a radar screen. Caffeine is available, however, in pill form and could be used as a performance enhancer. The major drawback of caffeine is its limited maximum effect. Particularly in people who a. eady use caffeine and have developed some tolerance, the boost in performance might not be large. Taken at certain times of the day, perhaps late at night when caffeine is not usually consumed, caffeine might significantly affect performance.

New Methylxanthines

The primary mechanism of action of methylxanthines such as caffeine is not yet proven. Methylxanthines inhibit the enzyme phosphodiesterase which degrades cyclic AMP. This potentiation of cyclic AMP was thought to be the mechanism of action for methylxanthines. However, recently a more potent action of methylxanthines, adenosine receptor blockade, is believed to be the more important effect of methylxanthines in the central nervous system. Since the known methylxanthines are so effective, research should be directed at determining (1) the sites of action of methylxanthines (phosphodiesterase, adenosine receptors, an unknown site?) and (2) new methylxanthines or analogs that share a similar site of action but may be more active than caffeine. This research may yield a militarily useful drug with caffeine-like actions but with increased potency. Possibly, a drug with no cross-tolerance with caffeine could be developed.

Vitamins

Vitamin supplements (e.g. "Stress Tabs") have been advertised to replace the vitamins depleted by exercise, stress or illness. There is no good evidence that vitamin intake in excess of normal requirements improves performance. Physicians often prescribe multivitamins for dieters or other patients who might not ingest sufficient food to provide daily requirements for vitamins. Perhaps soldiers in the field could take a multivitamin in case their food intake was inadequate. In rats, vitamin E deficiency diminishes exercise endurance capacity.

Ginsing and other substances

A variety of substances including Ginseng and bee products have been used as performance enhancers especially in non-US cultures, e.g. USSR and China. Scientific evidence for their effects on performance is lacking.

Discussion

Optimal nutrition, hydration, sleep and conditioning can improve physical and mental performance. The degree to which these measures can boost performance far exceeds the effects of currently available pharmacological aids. Training and conditioning increase blood supplies to exercised muscle, increase mitochondrial oxidative capacity to provide energy for work and improve the resistance of muscle to fatigue (1). Optimal nutrition provides adequate muscle and liver glycogen stores. Strategies of "carbohydrate loading" as practiced by marathon athletes might be useful to test scientifically to see whether endurance is really improved. Adequate water is

required to maintain performance especially when work generates sweat and water loss.

Many drugs are used by healthy normal civilians and soldiers. Stimulants are ingested daily as coffee, cola, tea, and cigarettes. Athletes continually are on the lookout for drugs that will improve performance. However, no effective currently available drug significantly increases speed, strength and endurance in rested healthy individuals. Some stimulants can increase alertness and delay sleep onset. These drugs have been used in special situations in past wars and will probably continued to be used. Improved training and conditioning, adequate hydration and nutrition, attention to sleep/rest cycles can significantly affect endurance and offer the best currently available methodology for optimizing human endurance at this time.

However, most scientific studies of performance enhancers have been performed on healthy rested individuals. It is possible that some of these drugs (methylxanthines, amphetamines, Ritalin) would significantly improve performance in a fatigued soldier. In addition, since fear can affect physical performance, anti-anxiety drugs might also counteract fatigue. Research into the effects of these stimulants and other drugs on performance in tired and/or stressed animals and man might produce useful guides for the use of these compound in military environments.

The potential usefulness of performance enhancing drugs should spur efforts to develop new compounds in this area.

References

1. Brooks, G.A. and Fahey, T.D. (1984) Exercise physiology New York, Wiley & Sons.
2. Costa, E. and Garattini, S., Editors. (1970) Amphetamines and related compounds New York, Raven Press,
3. Dews, P.B. Editor. (1984). Caffeine Berlin, Springer-Verlag.
4. Fowler, Gardner, G.W., and Egstrom, G.H. The effect of an anabolic steroid on physical performance of young men. (1965) J. Appl. Physiol. 20 1038-1040.
5. Jones, F.D. Sanctioned use of drugs in combat. (1985) In: P. Pichot, P. Berner, R. Wolf and K. Thau (editors) Psychiatry: the state of the art, vol VI, New York, Plenum Publishers.
6. Karpovitch, P.V. (1959) Effects of amphetamine sulfate on athletic performance. J.A.M.A. 170 118- 121.
7. Ritchie, J.M. (1975). Central nervous stimulants: the xanthines. In: L.S. Goodman and A. Gilman (editors) The pharmacological basis of therapeutics New York, Macmillan Publishing Co. pp 367-378.
8. Weiss, B. and Laties, V.G. (1962) Enhancement of human performance by caffeine and the amphetamines. Pharmacol. Rev. 14 1-36.

Pharmacological Optimization of Military Performance: Anti-Anxiety Agents

CPT J. M. Witkin and COL F. D. Jones
Division of Neuropsychiatry
Walter Reed Army Institute of Research
Washington, DC 20307-5100

The possibility of a "brave pill" was considered for potential field utility by the Psychopharmacology Study Group, Division of Neuropsychiatry, WRAIR. Such a compound might markedly enhance combat effectiveness through its direct psychomotor effects and indirectly through the prevention and treatment of psychological battle casualties. In addition to troop preparedness and effectiveness achieved through rigorous, realistic field training, use of these agents on the battle field may provide a decided tactical advantage. Although no such compound exists in pure form, drugs of several classes produce prominent behavioral effects which may be reasonable to consider at the present time.

The development of anxiety, chronic and disabling or acute and overwhelming, has generally been the most significant symptom related to combat breakdown during wars of the 20th century (Bar-On et al., 1983). Moreover, anxiety may significantly reduce combat efficacy in the remaining combat force (cf. Roberts, 1983). The nature of future battlefields is expected to rapidly produce anxiety and psychiatric casualties in excess of optimal levels. Moreover, the fluid, high-intensity, continuous warfare previously encountered (eg., 1973 Arab-Israeli War and 1982 War in Lebanon) revealed that the current doctrine of forward treatment (Glass, 1973) may be increasingly difficult to apply in future conflicts (Belenky et al., 1983). Such conditions of warfare not only underscore the need to maximize preventive measures but also call for novel approaches which include unit treatment of casualties during battle and even the use of prophylactic drugs. The prevention of excessive anxiety or its effective treatment after its appearance will have an important impact on combat strength and will undoubtedly be essential to success in future conflicts (Ingraham and Manning, 1980).

Fear and anxiety are related psychological phenomenon that share many common attributes with symptoms of apprehension and reduced ability to concentrate, perform, or make decisions. The causes of anxiety and fear are various but it is generally agreed that physiological and psychological stress, abundant in combat, is a principle instigator and maintainer of these states and an associated battle-shock syndrome (Belenky et al., 1983).

Our knowledge of the pharmacological basis of anxiety and fear has advanced tremendously in the past few years. Clinically-effective compounds (anxiolytics, anti-anxiety drugs, or minor tranquilizers) are presently available which may be of great field utility as prophylactics against anxiety and/or required tools for the treatment of these disorders once they arise. Benzodiazepines, so named for their chemical structure, are the most widely used anxiolytics today. Relative to previous generations of anti-anxiety agents, few deleterious side-effects or other pharmacological properties complicate the use of benzodiazepines.

Benzodiazepines produce four major effects: anxiolytic, anticonvulsant, muscle relaxant, and sedative. In addition to direct anxiolytic activity benzodiazepines may provide additional protection against stress; eg. these drugs have been reported to markedly reduce stress-induced ulcer formation. Benzodiazepines as a class vary widely in terms of their durations of action and the production of active metabolites. Although a host of benzodiazepine compounds have been described, no significant improvement in

activity over diazepam (Valium) or chlordiazepoxide (Librium) has been noted except for increases in the diversity of durations of action.

Earlier non-benzodiazepines, ethanol, barbiturates, and propanediols such as meprobamate and other more recent compounds such as methaqualone suffer from their lack of specificity of anxiolytic effect more so than do the benzodiazepines. In addition, these classic non-benzodiazepine sedative-hypnotic anxiolytic agents produce profound physical dependence (sometimes life-threatening) and psychological dependence as reflected in their abuse liability. Benzodiazepines suffer from these problems only to a limited extent.

Newly-developed non-benzodiazepine compounds show unique promise as anxiolytics. Zopiclone may be less sedating than benzodiazepines and tracazolate clearly shows a greater separation between anxiolytic and sedating doses. PK 8165 and PK 9080 are similar to tracazolate in this respect as are the triazolopyridazines (eg. CL 218,872) but these four compounds have yet to prove themselves in the clinic and to reliably show anxiolytic activity across a host of preclinical tests. All of these non-benzodiazepines appear to work through mechanisms much akin to those of the benzodiazepines (see below). The novel antipsychotic, clozapine, a dibenzodiazepine, also shows evidence of anti-anxiety potential.

Buspirone is perhaps the best available non-benzodiazepine for treatment of anxiety. It is non-sedating, has minimal effects on human psychomotor performance, does not seem to produce physical dependence and is unlikely to be a drug of abuse. In addition, buspirone counteracts catalepsy (immobility). Moreover, unlike other anxiolytics, buspirone does not interact with other central nervous system depressants such as ethanol to further impair performance. Buspirone is well advanced in clinical trials and may well take over much of the anxiolytic drug market. Although the mechanism of action of buspirone is unknown, it is clearly very different than that of the anxiolytics currently available. Elucidation of the mechanisms responsible for buspirone's effects will provide an important new development in our understanding of the neurobiology of anxiety and its control.

A behavioral effect shared by all anxiolytic drugs is the disinhibition of suppressed or punished behavior. The ability of anxiolytics to increase the occurrence of behavior suppressed by response-produced consequences may be the primary behavioral mechanism of anxiolysis. This pharmacological action serves as the basis for preclinical anti-anxiety drug screening. Performance enhancements are also noted with anxiolytics in situations in which behavior has been suppressed by a novel or stressful environment. Benzodiazepines not only increase already suppressed behavior but also protect against the suppressive effects of noxious stimulation (Witkin, unpublished observations).

Receptors in brain have recently been determined to be the primary structures which, by virtue of their specific binding of drug molecules, initiate the neural events associated with the anti-anxiety activity of benzodiazepines. Benzodiazepine receptors appear also to be involved in the mechanism of action of some non-benzodiazepine compounds such as the barbiturates. Multiple benzodiazepine receptor subtypes have been identified and this discovery may lead to the development of drugs which more precisely control anxiety. The existence of benzodiazepine receptors in brain suggest the involvement of a substance(s) inherent in brain which normally regulate neuronal substrates of anxiety. Several such endogenous ligands have been identified but no substance thus far has been definitively identified as the physiological regulator of anxiety. However, future studies in this area may eventuate in the establishment of novel and perhaps more beneficial means of controlling anxiety.

The binding of compounds to the benzodiazepine receptor initiates a complex series of neurochemical events. Two important systems have been studied more so than others: gamma-aminobutyric acid (GABA) and serotonin neurotransmission. Both of these systems appear to play important roles in the anxiolytic actions of benzodiazepines and other drugs; compounds which block serotonin receptors may also be useful anxiolytic

agents. The precise manner in which neurotransmission is affected by anti-anxiety agents is unknown.

Compounds have been developed in the past few years which can reverse (antagonize) or prevent all known effects of the benzodiazepines. Ro 15-1788 (flumazepil) is one such compound. Unlike Ro 15-1788, which selectively antagonizes effects of benzodiazepines and other compounds which directly bind to benzodiazepine receptors, CGS-8216 antagonizes pharmacological actions of some non-benzodiazepines as well (eg. barbiturates, meprobamate). Ro 15-1788 has little intrinsic activity but when given in relatively high doses it has some benzodiazepine-like actions. Another class of compounds, 3-substituted beta-carbolines, sometimes referred to as inverse agonists, may produce their benzodiazepine antagonist activity by producing pharmacological actions which are opposite to those of the benzodiazepines; these compounds can themselves induce wakefulness, anxiety, and seizure activity which is reversed by Ro 15-1788 or diazepam. Antagonists may soon be developed which will reverse all but select actions of antianxiety drugs. It must be recognized that the development of such antagonists raises the possibility of their use as offensive weapons. Such drugs might have important advantages over other incapacitating or lethal agents, particularly when used clandestinely.

Antagonists may allow for further refinement of the pharmacological effects of anxiolytic drugs. For example, in a combat setting, an anxious, hyperactive, possibly aggressive patient could be given a hypnotic dose of a benzodiazepine. In the later event of a requirement to move the patient to a forward treatment facility, the soldier could be given an antagonist to reverse all but the anti-anxiety actions of the benzodiazepine. For a soldier experiencing severe anxiety in a forward area, an anti-anxiety drug might be given by the medical aidman or even the squad leader. The soldier could then avoid all of the problems of rearward evacuation.

Another class of drugs that appears to posses anxiolytic activity is the beta-adrenergic antagonists such as propranolol and the alpha-adrenergic antagonists such as clonidine. Such agents could serve as a prophylactic treatment in soldiers exposed to battle conditions producing chronic autonomic arousal (Jones, 1983b). These agents seem to have little potential for dependence or abuse, and may be particularly effective when somatic elements of anxiety are prominent. Thus these compounds can improve weapon marksmanship by reducing normal muscle tremor which is accentuated by anxiety or fear. Normal components of the diet, amino-acids and neurotransmitter precursors, may produce psychological alerting effects as well as sedative or anxiolytic actions deserving of further examination (cf. Hegge and Tyner, 1983; Jones, 1983a).

The problems of drug tolerance, dependence, and withdrawal have severely limited the usefulness of sedative, tranquilizing agents in the past. Tolerance refers to the reduction in the effects of a drug that accompany repeated administration such that higher doses, with greater risks of side-effects, are required to reproduce initial drug effects. Tolerance, principally pharmacodynamic in nature, develops to many of the effects of the benzodiazepines such as sedation and muscle relaxation; anxiolytic activity of benzodiazepines, however, does not diminish with repeated dosing but may increase. In fact, a "drug-naive" effect has been reported in which initial treatment does not result in anti-anxiety effects. Thus some of undesirable effects of the benzodiazepines such as decrements in motor and cognitive function may be mitigated by repeated administration.

Drug dependence and withdrawal may present more serious problems in combat. Two types of dependence may develop upon repeated drug administration. Psychological dependence refers to the establishment of a drug as a positive reinforcer; dependence of this sort is associated with the maintenance of behavior by the acquisition and self-administration of the drug and is typically accompanied by feelings of drug craving. Physical dependence is defined by the production of physiological and behavioral changes (withdrawal signs) which occur upon the abrupt termination of chronic drug

intake or upon administration of an antagonist. Although physical dependence need not be present for psychological dependence to develop, these two types of dependence severely compromise and complicate the use of sedative-hypnotic/anxiolytic drugs in the field. Physical dependence with rebound anxiety, autonomic arousal, and insomnia have been documented for the benzodiazepines as well as for earlier generation sedative-hypnotic/anxiolytics such as the barbiturates which produce additional withdrawal symptoms (eg. convulsions, hallucinations, death). Shorter acting benzodiazepines appear to have greater physical dependence liability than longer acting ones. Of the antianxiety agents currently in use, the benzodiazepines are least abused, and least likely to produce dependence problems. Newer anxiolytics (eg. buspirone) may be completely free of these problems.

Concurrent use of antianxiety agents with other compounds may result in unwanted effects and may in some cases produce life-threatening consequences. Currently used antianxiety agents suffer from their prominent interactions with other central nervous system depressant drugs including other sedative compounds and ethanol resulting in sedation, hypnosis, and ultimately respiratory arrest. Newer anxiolytics appear to be relatively immune to this synergistic action. In other cases, combinations of anxiolytics with other drugs produce some desirable outcomes. When given in conjunction with appropriate doses of d-amphetamine, benzodiazepines and barbiturates may produce greater anxiety reduction than either drug alone as well as a qualitatively unique pharmacological profile. Although such drug combinations may be quite prone to drug abuse, their unique behavioral and subjective effects may render them more efficacious than anxiolytics alone as brave pills (eg., feelings of invulnerability). Drug combinations may also allow use of lower levels of the anxiolytic drug without compromising the clinical or field objective. For example, thyrotropin-releasing hormone potentiates the antianxiety actions of benzodiazepines, barbiturates and ethanol, without having actions of its own; thus the dosage levels of the sedative/anxiolytic agents can be reduced, lessening chances for deleterious side-effects (Witkin et al., 1984). Other clinically-important or militarily-relevant adverse benzodiazepine-drug interactions are rare.

All drugs produce unwanted effects although some compounds do so to a greater extent than others. These so-called side-effects, discussed above, must be considered. For example, animal studies have indicated that anxiolytics may decrease aggression. The consequences of possible disruptions in normal combat-stress reactions resulting from the use of drugs is not well defined. Drugs do not produce unitary effects on behavior; behavioral effects of drugs depend upon a host of subtle controlling variables such as environmental context, motivation, and behavioral and pharmacological history (Barrett and Witkin, 1985; McKearney and Barrett, 1976). Interaction of drugs and behavior on the battlefield can not be precisely predicted. Soldier compliance with a prophylactic drug doctrine may also be a problem. Some soldiers may benefit more than others from such treatment whereas some personnel may be treatment risks (eg. potential drug abusers). Means of identification of these subpopulations are desirable.

Drugs are currently available which relieve anxiety yet produce few deleterious side-effects. The benzodiazepines and buspirone look most promising in this regard. The development and refinement of novel drugs for controlling fear and anxiety will likely be forthcoming. The availability of antagonists to the anxiolytic agents provides a rapid means of controlling drug action; eg. reversal of drug overdose or idiosyncratic effects. Non-drug techniques for controlling fear and anxiety (particularly diet, behavioral conditioning and stress- inoculation) may also be useful alone or in combination with anxiolytic agents. Rigorous laboratory and field studies are urgently required to evaluate the utility of these ideas. The technical skills, physical conditioning, confidence, unit cohesion and esprit, natural by-products of good training and leadership, are key components of an effective fighting force. None of the drugs discussed above can substitute for appropriate military training and leadership but may provide decided

combat advantages with appropriate field use.

References

Barrett, J. E. and Witkin, J. M. (1985). The role of behavioral and pharmacological history in determining the effects of abused drugs. In S. R. Goldberg and I. A. Stolerman, eds., Behavioral Analysis of Drug Dependence, Academic Press, New York.

Bar-On, R., Soloman, Z., Noy, S., and Nardi, C. (1983). War-related stress: an overview of the clinical picture in the 1982 conflict in Lebanon. Paper presented at the third International Conference on Psychological Stress and Adjustment in Time of War and Peace, Tel Aviv, Israel, 2-6 January, 1983.

Belenky, G. L. and Tyner, C. F., and Sodetz, F. J. (1983). Israeli Battle Shock Casualties: 1973 and 1982. Report WRAIR NP -83-4, Walter Reed Army Institute of Research, Washington, DC.

Glass, A. J. (1973). Lessons learned. In A. J. Glass, ed., Neuropsychiatry in World War II, Volume II, Overseas Theaters, Government Printing Office, Washington, DC.

Hegge, F. W. and Tyner, C. F. (1982). Deployment Threats to Rapid Deployment Forces. Division of Neuropsychiatry report, Walter Reed Army Institute of Research, Washington, DC.

Ingraham, L. H. and Manning, F. J. (1980). Psychiatric battle casualties: The missing column in a war without replacements. Military Review, August 1980, 19-29.

Jones, F. D. (1982). Combat stress: Tripartite model. International Review of the Army, Navy, and Air Force Medical Services, 55: 247-254.

Jones, F. D. (1983a). Sanctioned use of drugs in combat. In P. Berner, ed., Psychiatry: The State of the Art, Volume VI, Plenum, New York.

Jones, F. D. (1983b). Combat psychiatry in modern warfare. In O. Adelafa and F. D. Jones, eds., War and Its Aftermath, John West, Nigeria.

McKearney J. W. and Barrett, J. E. (1978). Schedule-controlled behavior and the effects of drugs. In D. E. Blackman and D. J. Sanger, eds., Contemporary Research in Behavioral Pharmacology, Plenum Press, New York, 1-68.

Roberts, L. (1983). Anxiety state and battlefield performance: a negative correlation. AST-266OR-031-83, May, 1983.

Witkin, J. M., Sickle, J. and Barrett, J. E. (1984). Potentiation of the behavioral effects of pentobarbital, chlordiazepoxide and ethanol by thyrotropin-releasing hormone. Peptides 5: 809-813.

Pharmacological Optimization Of Performance: Sleep and Arousal

CPT Vincent M. O'Donnell

Division of Neuropsychiatry

Walter Reed Army Institute of Research

Washington, D.C. 20307-5100

Abstract

Numerous hypnotic (sleep-inducing) drugs are approved and available in the United States. Of these drugs, the benzodiazepine hypnotics appear to be the drugs of choice for military operations due to their relatively minor performance effects as compared to other hypnotic compounds. Important differences exist, however, between the military and civilian applications of these drugs in terms of both efficacy and effects on human performance. Additional research is needed to answer questions raised by these application differences. The initial military use of hypnotic compounds may be most appropriate to long-distance air deployment of troops where there is a reasonable probability that personnel will be able to sleep for six or more hours subsequent to drug administration. Additional military uses for hypnotics may be dependent upon the fielding of a selective counteragent to the hypnotic compound.

Military operations often require that personnel work for several days with little or no sleep. When opportunities for sleep do occur, they are usually quite brief and are not necessarily in synchronization with the individual's normal sleep schedule. Further, the physical and psychological conditions of military operations are far from the dark, quiet, secure, relaxed, and comfortable environment that would promote the rapid onset and maintenance of restful and restorative sleep. Hypnotics, drugs which speed the onset of sleep and help to maintain sleep, are therefore of potential value in military operations.

Because there is a large market for hypnotic drugs in clinical practice (e.g. in cases of insomnia and other types of sleep pathology), there are many approved hypnotic compounds on the market today. Hypnotics available for use in the United States include various anti-histamines and barbiturates, chloral hydrate and chloral betaine, ethchlorvynol, ethinamate, methprylon, and the benzodiazepines. In comparison with the benzodiazepines, the other listed drugs have serious adverse physical and behavioral effects, and some have a very high abuse potential. As a result, the benzodiazepines (e.g. triazolam, temazepam, and flurazepam) have become the most frequently prescribed hypnotics in this country.

Recent military operations have made use of commercially available hypnotics. Israeli forces, for example, administered a commercially available compound to promote

sleep among their troops during the air deployment portion of the Entebbe raid. Temazepam was used successfully by British pilots during the Falklands conflict to enable the best use of irregularly scheduled and infrequent sleep periods under uncomfortable conditions. While the availability of numerous approved hypnotic compounds may encourage their use in military operations, it is important that the substantial differences between military and civilian applications of these drugs be fully understood.

Perhaps the most important of these differences is the greater impact of drug-associated performance decrements on military as opposed to civilian applications. While the benzodiazepines are the class of currently available hypnotics associated with the lowest incidence and magnitude of adverse effects, these effects are still sufficient to cause problems in military operations. The most obvious performance effect of benzodiazepines and other hypnotic compounds is, of course, the induction of sleep. The concomitant reduction in arousal and alertness will affect virtually all types of performance, especially vigilance and sustained attention tasks. In addition, the benzodiazepines have several effects on performance which may differ etiologically from their sleep-inducing effect. These drugs have been demonstrated to impair memory function, especially the consolidation of newly learned material into long-term memory (Clarke, Eccersley, Frisby, and Thornton, 1970; Lister, 1985). A question has been raised as to whether this memory deficit is a primary effect of the drug or whether it is a secondary effect resulting from the drug's hypnotic action--i.e., the drug puts the subject to sleep and in so doing, prevents memory consolidation from occurring (Roehrs, McLenaghan, A., Koshorek, G., Zorick, F., and Roth, 1984). The existence of studies which show memory impairment in subjects who are kept awake after benzodiazepine administration would argue against this theory (Ghoneim and Mewaldt, 1975, 1977). Speed of performance is also consistently slowed by benzodiazepines so that any task involving time pressure (e.g. reaction time, symbol copying) is substantially decremented (Johnson and Chernick, 1982).

Because such effects are of some concern in virtually any population, an effort has been made to select as hypnotics those benzodiazepines which have a relatively short plasma half-life. In civilian use, the chosen hypnotic would hopefully produce a rapid and relatively brief hypnotic effect, during which the subject would refrain from certain critical activities such as driving and the operation of complex machinery. In the military context, the problem is not so simple and the solution not so efficacious.

First, it should be noted that the plasma half-life of a benzodiazepine does not give a particularly strong indication of its onset and duration of actions. While it is true that, in general, the long half-life benzodiazepines have a longer period of action than the very short half-life benzodiazepines, there are some exceptions. Clobazam, for example, is metabolized like a long half-life benzodiazepine, but has a time course of behavioral effects more like a short half-life benzodiazepine. Further, within a given category of benzodiazepines (e.g. short half-life), their is only a loose correlation between half-life and duration of behavioral effects (Hindmarch, Ott, and Roth, 1984).

Next, the duration of behavioral effects is likely to be considerably longer than we would hope for in military operations even with very short-acting benzodiazepines. Triazolam, for example, has a half-life of approximately 3.8 hours. The behavioral effects of a 0.5 mg dose of the drug (a normal clinical dose) have been demonstrated to last at least 6 hours and perhaps as much as 10 hours following administration of the drug (Nicholson and Stone, 1980; Veldkamp, Straw, Metzler, and Demissianos, 1974; Roth, Hartse, Saab, Piccione, and Kramer, 1980). The duration of these behavioral effects is of considerable importance given that there are few situations in military operations when

the users of the hypnotic drug can consistently restrict their activities to those not affected by the hypnotic for such an extended period. Further, the consequences of impaired performance in a military operation are likely to be graver than in a civilian situation. One of the few military situations where personnel have a good chance of remaining inactive for a predictable and extended length of time is in long-distance air deployment, since trans-oceanic flights of six to 10 hours are common.

While the use of hypnotics in such troop deployments would seem appropriate, there are additional considerations. First, the efficacy of benzodiazepine and other hypnotics has not been established under conditions similar to those of military deployment operations. In clinical efficacy studies, an environment is provided which is extremely conducive to sleep, and the drug is administered around the normal sleep time. Contrast this with the environment of the troop transport operation. In daytime operations, the plane may be well-lit and the drug may be administered at a high point in the circadian rhythm. The plane will be noisy and is often quite hot or quite cold. The troops will have a high level of anxiety and autonomic arousal related to their mission, and will be forced to sit upright in crowded conditions for an extended period. While no clinical efficacy studies have come close to duplicating this environment, some facets of it have been studied individually in relation to hypnotic efficacy. Triazolam, for example, has been shown to produce sleep in subjects about to undergo surgery, who probably have high levels of anxiety and autonomic arousal (Keighley, Gannon, Warlow, Jenkins, and Gammon, 1980). The same drug has been demonstrated to induce sleep or drowsiness at times other than night (Seidel, Roth, Roehrs, Zorick, and Dement, 1984; Walsh, Muehlbach, and Schweitzer, 1984) and in subjects who were not kept in typical sleep environments (Gorenstein and Gentil, 1983). There is thus evidence that some hypnotics may be efficacious in the environment of military air deployment operations, but further research is needed.

Because most clinical uses of benzodiazepine hypnotics involve repeated bedtime administrations, there is little available information on possible delayed effects of a single administration. During the minor circadian trough which occurs during the day, a reduction in performance levels typically occurs. This reduction appears to be aggravated by a previous night of benzodiazepine-induced sleep (Hindmarch, et al., 1984). Further, EEG changes at 16.5 to 17.5 hours after administration of a therapeutic dose of triazolam have been demonstrated, well after plasma levels are virtually undetectable (Veldkamp, et al., 1974). What a single administration of a hypnotic at a high point in the circadian rhythm will do to subsequent natural sleep periods is unknown.

The use of benzodiazepine hypnotics in military operations would therefore appear most appropriate to long-distance air deployment of troops. Even when hypnotic use is restricted to such operations, there are questions involving the efficacy and performance effects of hypnotic drug use. While the efficacy question can be resolved by further research, the problems of adverse performance effects may require specific countermeasures. Obviously, it cannot be guaranteed that troops can count on six to eight hours of uninterrupted sleep even during deployment operations. Emergency landings, briefings, assignment changes may all require a high level of alertness and performance. Further, it is the leaders who are most likely to be awakened in the midst of air deployment for changes of plans and decisions. This involves something of a paradox since troop leaders perform the tasks that are most sensitive to sleep deprivation and, at the same time, are the personnel least likely to have adequate time for sleep. Thus the personnel who could most benefit from the use of a hypnotic compound are those who would be most impacted by its performance effects and therefore might be most resistant to its use.

The use of a drug which would specifically counteract the sedating and performance effects of a hypnotic compound could solve these problems by safely awakening personnel and rapidly bringing them to full performance levels. Such a counteragent drug would also expand the potential use of hypnotics. If troops may be rapidly and reliably awakened, then it would not be necessary to restrict the use of hypnotics to occasions when action is not likely to be required for six to eight hours. Ground troops, pilots, and shipboard personnel who will be off duty for only a few hours could use a hypnotic and its counteragent to make optimal use of their rest periods.

This strategy will, however, require a **selective** counteragent; a general stimulant may awaken the subjects, but will carry with it its own potential for adverse performance effects and may not counter all of the negative effects of the hypnotic. In this respect, the benzodiazepines have an additional advantage over other compounds. While no selective benzodiazepine antagonist has yet received approval for human use in this country, several have been used in experimental studies and hold promise of eventual clinical use (Loew, Nienow, Lawson, Toll, and Uyeno, 1985; Haefely, Bonetti, Burkard, Cumin, Laurent, Mohler, Pieri, Polc, Richards, Schaffner, and Scherschlicht, 1983). On the other hand, many alternate hypnotics such as the barbiturates and ethinamate have either unknown or extremely diverse mechanisms of action, making the development of a selective counteragent unlikely.

In addition to the classes of drugs cited above, experimental work with several endogenous substances such as interleukin-1, delta sleep-inducing peptide, and substance S hold hopes for the future development of hypnotics which may be safer and more efficacious than those currently available (see for example, Wauquier, Gaillard, Monti, and Radulovacki, 1985). As with current compounds, novel hypnotics will need to be evaluated against the unique requirements of military operations and should have a selective and rapidly acting antagonist if they are to receive maximum military application.

REFERENCES

Clarke, P.R.F., Eccersley, P.S., Frisby, J.P., Thornton, J.A. (1970) The amnesic effect of diazepam (Valium). British Journal Of Anesthesiology, 42, 690-697.

Ghoneim, M.M. and Mewaldt, S.P. (1975) Effects of diazepam and scopolamine on storage, retrieval and organizational processes in memory. Psychopharmacologia, 44, 257-262.

Ghoneim, M.M. and Mewaldt, S.P. Studies on human memory: the interactions of diazepam, scopolamine, and physostigmine. (1977) Psychopharmacology, 52, 1-6.

Gorenstein, C. and Gentil, V. (1983). Residual and acute effects of flurazepam and triazolam in normal subjects. Psychopharmacology, 80, 376-379.

Haefely, W., Bonetti, E.P., Burkard, W.P., Cumin, R., Laurent, J.-P., Mohler, H., Pieri, L., Polc, P., Richards, J.G., Schaffner, R. and Scherschlicht, R. (1983) Benzodiazepine antagonists. In E. Costa (Ed.), Benzodiazepines--From Molecular Biology To Clinical Practice, New York: Raven Press, pp. 137-146.

Hindmarch, I., Ott, H., and Roth, T.(Eds.) (1984) Psychopharmacology (Supplementum I). Berlin: Springer-Verlag.

Johnson, L.C. and Chernik, D.A. (1982). Sedative hypnotics and human performance. Psychopharmacology, 76, 101-113.

Keighley, M.R.B., Gannon, M., Warlow, J., Jenkins, C.R.M., and Gammon, R.J. (1980). Evaluation of single-dose hypnotic treatment before elective operation. British Medical Journal, 281, 829-831.

Lister, R.G. (1985) The amnesic action of benzodiazepines in man. Neuroscience and Biobehavioral Review, 9, 87-94.

Loew, G.H., Nienow, J., Lawson, J.A., Toll, L., and Uyeno, E.T. (1985) Theoretical structure-activity studies of beta-carboline analogs, Requirements for benzodiazepine receptor affinity and antagonist activity. Molecular Pharmacology, 28, 17-31.

Nicholson, A.N. and Stone, B.M. (1980). Activity of thehypnotics, flunitrazepam and triazolam, in man. British Journal of Clinical Pharamacology, 9, 187-194.

Roehrs, T., McLenaghan, A., Koshorek, G., Zorick, F., and Roth, T. (1984) Amnesic effects of lormetazepam. Psychopharmacology, 1, 165-172.

Roth, T., Hartse, K.M., Saab, P.G., Piccione, P.M., and Kramer, M. (1980). The effects of flurazepam, lorazepam, and triazolam on sleep and memory. Psychopharmacologica, 70, 231-237.

Seidel, W.F., Roth, T., Roehrs, T., Zorick, F., and Dement, W.C. (1984). Treatment of a 12 hour shift of sleep schedule with benzodiazepines. Science, 224, 1262-1264.

Veldkamp, W., Straw, R.N., Metzler, C.M., and Demissianos, H.V. (1974). Efficacy and residual effect evaluation of a new hypnotic, triazolam. The Journal of Clinical Pharmacology, 14, 102-111.

Walsh, J.K., Muehlbach, M.J. and Schweitzer, P.K. (1984). Acute administration of triazolam for the daytime sleep of rotating shift workers. Sleep, 7, 223-229.

Wauquier, A., Gaillard, J.M., Monti, J.M., Radulovacki, M. (Eds.) (1985) Sleep Neurotransmitters and Neuromodulators. New York: Raven Press.

Pharmacological Optimization of Performance: Memory and Cognition

CPT John R. Leu
LTC Frederick J. Manning
CPT Vincent M. O'Donnell

Division of Neuropsychiatry
Walter Reed Army Institute of Research

Abstract

Decrements in reasoning, vigilance, and memory will surely occur under extremely stressful circumstances. Recent advances in the neurochemistry and neurophysiology of cognitive processes suggest that it may be possible, in the foreseeable future, to improve or maintain these abilities under severe conditions. No currently available compounds have been shown to reliably and unambiguously improve cognitive functioning in normal rested individuals. A few compounds show some evidence of effectiveness in aged or dysfunctional individuals. These and other treatments should be examined for effectiveness in normal but severely challenged individuals. Emphasis should be on attenuating the performance deficits that result from stressful situations rather than on improving the performance of normal individuals under optimal circumstances.

Introduction

> "This is One-six. I'm at coordinates 298028, over." "This is Delta Six. Are you *sure of that? . . . According to what you gave me, you're in the South China* Sea, over." . . . Goddamn! I've done this a million times . . . "Concentrate", I said to myself while staring at the meaningless blob of colors, whirls, and numbers. Suddenly, everything jelled in perfect clarity. "Delta Six, this is One-six. You're right, my coordinates are 288028. I need a dustoff fast, over." . . . Two lessons: claymores and thunderstorms do not mix, and reading a map under pressure is not like reading one in a training exercise. (Downs, 1978).

Lieutenant Downs learned the hard way that even well practiced mental tasks can suddenly become confusing or even impossible under extreme conditions. Hunger, physical and mental fatigue, fear, pain, hypoxia, intoxication and many other stressors can reduce cognitive efficiency. Memory, ability to concentrate, and reasoning ability can all be drastically impaired under severe circumstances. Often, the consequences of such changes are worsened by occurring without being obvious to the affected individual. With the incredible complexity of a modern battlefield soldiers now are faced with mental tasks of enormous difficulty. Advances in the technology of warfare dictate that intense battles can now continue during darkness, bad weather and other conditions that previously would have brought about at least a lull in the battle. Both armaments and doctrine now call for soldiers to be engaged in sustained operations lasting as long as several days without significant rest. The missing element in these scenerios is likely to be the inability of even well trained, highly motivated soldiers to satisfactorily perform their assigned tasks for such extended periods. This paper examines issues involved in and approaches to enhancement of cognitive performance under challenging conditions.

The search for ways to improve performance is not new. Recent advances in neuroscience have however resulted in many new approaches. These same advances have resulted in an increased awareness of the extent to which neural functions do not operate independently but are quite interrelated. It is, therefore, important that we carefully

consider the trade-offs when seeking to optimize any one aspect of performance.

The analysis of cognitive functioning can be considered to have three essential phases: First a training phase where some new skill or information is acquired; second, a delay period; and finally a testing phase. For the purposes of studying performance during sustained operations we would consider the delay period to be disruptive. Interventions might be effective during any or all of these phases. Treatments are evaluated by the degree to which they reverse or attenuate the disruption in performance usually seen under the test conditions. Most animal research screening for cognitive enhancing agents uses single day paradigms. A typical procedure is for animals to be trained on an avoidance task in the morning. These animals are then exposed to an intervention which would normally lower subsequent avoidance performance. Such interventions are most often electrically induced seizures or hypoxia. The effectiveness of various treatments in minimizing this induced loss of performance is taken as an indication of their possible effectiveness (Gamzu, 1984). Human research normally uses persons who have suffered serious declines in mental abilities due to normal aging or disease. Seldom does research concerning enhancement of cognitive functioning use normal, well trained subjects (animal or human) performing under sub-optimal conditions. Proper evaluation of any potential interventions for military use during sustained operations must include such tests.

Potential Performance Enhancers: Memory and Cognitive Functioning

Nootropics:

These recently developed "mind-activating" drugs are purported to have selective activity on telencephalic structures and to have a very low incidence of toxicity and undesired effects. The mechanism of action is not well understood but may work at least in part by improving the ATP/ADP ratio. A large number of related compounds have been synthesized and several of these drugs are currently in use (in Europe) for the treament of various dementias. Results of these clinical studies have been mixed, but there have been enough positive findings to encourage continued research. A few studies have even demonstrated small but reliable improvements in cognitive performance in normal young adults. Piracetam, a prototypical example of the nootropic class, has been shown to improve mental performance in aged but clinically normal individuals (Mindus, et al, 1976) and in young adults (Dimond & Brouwers, 1976).

Stimulants:

Caffeine, and to a lesser extent in recent years nicotine, have been fixtures of life in the U.S. Army for decades. Coffee has in fact been a staple of Army rations since 1832. Though widely viewed as a mild stimulant helpful in temporarily relieving minor fatigue and boredom with little risk, laboratory studies have reported improvements in tasks ranging from vigilence and reaction time to complex verbal tasks like the Graduate Record Exam (Sawyer, Julia, and Turin, 1982), although effects on the more purely cognitive tasks seem to be dependent upon personality. Nicotine, in tablets as well as cigarettes, has repeatedly been shown to help sustain performance on monotonous tasks, and to improve both speed and accuracy on a variety of information processing tasks (Wesnes and Warburton, 1983). It is clear that, given the social acceptibility of these drugs, any proposed stimulant will have to show marked advantages over these two to merit much consideration.

The heaviest reported use of other stimulants is the extensive use of amphetamines by the Soviet Red Army during World War II. Reported benefits included

not only reduction of fatigue and drowsiness but also improvements in memory and concentration (Jones, 1985). Amphetamine and related stimulants (methylphenidate -- Ritalin) do in fact improve performance in routine tasks requiring vigilance. The tradeoff is perseverative stereotypic behavior. Complex reasoning and decision making will likely be adversely affected.

Hormones:

Several investigators, including Kety (1972), McGaugh (1973, 1983) and de Weid (1974) have implicated a role for endogenous hormones as modulators of memory processes. In particular, catecholamines and endorphins seem to effect memory storage. Adrenocorticotropic hormone (ACTH) has also been shown to modulate memory storage processes. All of these systems are activated during arousal and so would be especially important when considering cognitive functioning under challenging circumstances. Although there is usually a "U" shaped function to the effects, moderate increases in ACTH, increases in catecholamines, and blockade of endorphins have all been shown to enhance memory in laboratory animals. The main lesson from this work might be that there is an optimal level of arousal for a given behavior at a specific stage of training. Too much or too little arousal will result in lower performance.

Enhancers of Neurotransmitter Systems:

The discovery of decreased levels of cholinergic activity in Alzheimier patients prompted a great deal of research into the effectiveness of treatments which would increase cholinergic functioning. There are two basic permutations of this approach. The first approach is to increase the levels of acetylcholine (ACH) in the brain by increased systhesis. Various dietary precusors such as choline and lecithin have been tried with minor effects at best. The other permutation involves more direct manipulation of the cholinergic system, either by direct use of agonists or by using cholinesterase inhibitors to prolong the action of what ACH is available (Bartus, et al, 1984). Physostigmine and other drugs increasing the duration of action of ACH have been shown to improve memory in laboratory animals,and, under certain conditions, in humans. Non-specific neural excitants such as strychnine have also been found to improve memory under certain conditions (Crabbe & Alpern, 1973).

Blood Flow/ Metabolism Enhancers:

Under some severe conditions an increase in the amount of blood flow to the brain might improve performance. Several types of drugs are available for accomplishing this. Pentoxifylline (Trental) increases the flexibility of blood cells so that more blood can get into areas having low blood flow (Petrie, 1985). Compounds of this type are suggested to be effective against the primary symptoms of senility such as disorientation, confusion and memory loss. Their effectiveness in normal individuals under stressful conditions should be evaluated. Calcium channel blockers (e.g. Nifedipine, Verpapamil, Diltiazem) can prevent spasms in muscles and blood vessels. This type of drug might be useful in maintaining effective cognitive functioning at high altitudes, in aircraft with low-cabin pressures or other circumstances leading to brain oxygen debt.

Anxiolytics:

Beta-adrenergic blockers (e.g. Propranolol) have proven useful for reducing performance anxiety in concert musicians and actors as well as competitive shooters (Noyes, 1985). These might improve cognitive performance under the stressful conditions of combat as well.

RNA Synthesis:

Increasing rates of RNA synthesis was thought to be a mechanism for improving memory. Magnesium Pemoline was intially purported to increase RNA synthesis and a number of studies followed investigating its effectiveness in improving memory. After the initial unbridled enthusiasm it has become evident that the effects are small if they exist at all (Eisenstein & D'Amato, 1975).

Future Directions

Our laboratories at Walter Reed are currently conducting a number of research projects concerned with optimization of cognitive functioning. Animal studies include projects investigating: performance decrements in rats living in a chronic stressful environment; performance changes occurring under the influence of chemical warfare antidote compounds; and circadian patterns of cognitive performance. Human studies include: ongoing studies concerned with performance changes under the influence of chemical warfare antidote compounds; continuing studies investigating performance changes with sleep loss; and planned studies of the effects of nootropic and alerting agents on performance changes accompanying prolonged sleep loss.

Summary

* When the going gets tough, anybody can get stupid.
* Naturally occuring hormones, including ACTH, epinepherine, and endorphins, play an important role in memory acquisition, consolidation and retrieval.
* Training should be under conditions of physiological and emotional arousal similar to those under which the tasks are likely to be performed.
* Consideration must be given to performance variables: These include: attention and perception, motivation, degree and type of required repsonse, and task complexity.
* Although a few reports have shown minor performance improvements in normal individuals, there are no known pharmacologic compounds that are generally regarded as be useful for improving congnitive functioning in rested, normal individuals. A number of agents have shown some effectiveness in improving functioning in aged or dysfunctional individuals; their effectiveness in improving degraded performance in normals should be systematically studied.

References

Bartus, R. T., Dean, R. L., Pontecorvo, M. J., & Flicker, C. (1984) The cholinergic hypothesis: A historical overview, current perspective and future directions. New York Academy of Sciences Conference on Memory Dysfunctions.

Crabbe, J. C. & Alpern, H. P. (1973) Facilitation and disruption of the long-term store of memory with neural excitants. Pharmacology, Biochemistry, and Behavior, 1, 197-202.

de Weid, D. (1974) Pitutitary-adrenal system hormones and behavior. The Neurosciences Third Study Program (F. O. Schmitt and F. G. Worden, eds.). MIT

Press, Cambridge, Mass, pp 653-666.
Dimond, S. J. & Brouwers, E. Y. M. (1976). Increase in the power of human memory in normal man through the use of drugs. Psychopharmacology.
Downs, Frederick. (1978). The Killing Zone. Berkley Books, New York City, New York. Pages 42-44.
Eisenstein, N. & D'Amato, M. R. (1975) Effects of magnesium pemoline on a delayed match-to-sample task in monkeys. Behavioral Biology, 15, 245-250.
Gamzu, E. (1984). The effects of aniracetam (Ro 13-5057) on the enhancement and protection of memory in mice. New York Academy of Sciences Conference on Memory Dysfunctions.
Jones, F. D. (1985). Sanctioned use of drugs in combat. Psychiatry: The State of the Art. Vol VI, Plenum, New York.
Kety, S. S. (1972) Brain catecholamines, affective states, and memory. The Chemistry of Mood, Motivation and Memory. (J. L. McGaugh, ed). Plenum, New York. pp 65-80.
McGaugh, J. L. (1973) Drug facilitation of learning and memory. Annual Review of Pharmacology, 13, 229-241.
McGaugh, J. L. (1983) Hormonal influences on memory. Annual Review of Psychology, 34, 297-323.
Mindus, P., Cronholm, Levander, S. E., & Schalling, D. (1976). Acta Psychiat. Scand., 54, 150-160.
Noyes, R. (1985) Beta-adrenergic blocking drugs in anxiety and stress. Psychiatric Clinics of North America, 8, 119-132.
Petrie, W. M. (1985, September). Quoted in American Health.
Sawyer, D.A., Julia, H.L., & Turin, A.C. (1982) Caffeine and human behavior: arousal, anxiety, and performance effects. Journal of Behavioral Medicine, 5, 415-439.
Wesnes, K. & Warburton,D.M. (1983) Smoking, nicotine, and human performance. Pharmacological Therapeutics, 21, 189-208.

Appropriateness of the Armed Services Vocational Aptitude Battery for Ninth and Tenth Graders

Toni G. Wegner
Air Force Human Resources Laboratory

Abstract

Data were collected on 15,252 ninth through twelfth graders to examine issues in the development of norms on the Armed Services Vocational Aptitude Battery (ASVAB) for ninth and tenth graders. The recently implemented reference population of the ASVAB, identified to be representative of 1980 American youth, currently provides norms for enlistment and normative information for eleventh and twelfth graders who take the ASVAB. Ninth and tenth grade norms are desired to provide valuable information to these students and their counselors and to enhance the recruiting market. The goal of the present research was to examine the appropriateness of the ASVAB for ninth and tenth graders by looking at score distributions, item difficulty levels, reliabilities, and numbers of omits across ninth through twelfth graders. Results suggest ninth and tenth grade ASVAB scores provide similar information to eleventh and twelfth grade scores and that the ASVAB can be useful for ninth and tenth graders.

Introduction

The ASVAB is an aptitude battery used for the selection and classification of all military enlisted personnel. It is administered annually to about one million service applicants through the operational testing program, and to about one million high school students through the Department of Defense (DOD) Student Testing Program. In addition to providing information to high school students and counselors, the Student Testing Program is a major recruiting tool used to identify prospective enlistees. It is targeted at eleventh and twelfth graders because their scores can be used for enlistment. Schools participate in this program on a voluntary basis, and it has been acceptable in the past to test tenth graders in order to encourage schools to participate. Although ninth graders have been tested on the ASVAB in the past, this testing is no longer permitted.

When students test on the ASVAB in schools, their scores are reported in percentiles based on the military reference population and also in terms of grade and sex norms. ASVAB Form 14, currently used for testing in the schools, provides norms for eleventh and twelfth graders based on a representative sample of 1980 American youth (Maier & Sims, 1982); however, norms are not available for ninth or tenth graders. This is because the ASVAB was designed for persons of enlistment age (18-23 years). Although there is a push by recruiters for ninth and tenth grade norms, an issue that must be addressed first is whether the ASVAB is appropriate for ninth and tenth graders. That is the focus of this paper.

Sample

Data were collected as part of the DOD Student Testing Program from 15,252 ninth through twelfth graders in high schools across the country. Participation from schools was voluntary, but schools that participated had to agree to test all ninth through twelfth graders in attendance on the day of testing. Testing was conducted between September 1984 and January 1985.

Data editing was done to eliminate cases with bad demographic or school codes. In addition, six percent of the cases were eliminated because school attendance figures indicated school participation was limited. Another two percent of the cases were deleted because one or more subtests contained all omits.

The final sample contained 14,061 students (51% male, 49% female) from 47 schools. School size ranged from less than 50 to greater than 1,000, with the majority of schools testing between 100 and 500 students. The sample contained 18% Blacks, 10% Hispanics, and 72% other (including Whites). By geographic region, 27% were from the Northeast, 26% from the North Central, 28% from the South, and 19% from the West. Compared with 1980 Census figures, minorities were overrepresented and people from the South were underrepresented.

Method and Results

Comparisons were made between ninth, tenth, eleventh, and twelfth graders to look at differences in performance on the ASVAB between grades. Analyses were conducted on the eight power subtests of the ASVAB: General Science (GS), Arithmetic Reasoning (AR), Word Knowledge (WK), Paragraph Comprehension (PC), Automotive and Shop Information (AS), Mathematics Knowledge (MK), Mechanical Comprehension (MC), and Electronics Information (EI). The two speeded subtests, Numerical Operations (NO) and Coding Speed (CS), were analyzed differently and are not included here because of space limitations. Analyses included comparisons of score distributions, item difficulty levels (including items with chance difficulties), and omitted items.

A comparison of mean subtest scores for the four grades revealed a generally consistent trend across the subtests. Mean scores consistently increased with increases in grade with about one and one-half raw score points difference between ninth and tenth grades and the same between tenth and eleventh grades. The difference between eleventh and twelfth grade mean scores was only about one-half to three-fourths raw score points. For example, mean scores for ninth through twelfth graders on Arithmetic Reasoning were 13.23, 14.68, 16.44, and 17.28, respectively. The score distributions reflected the mean differences, but were generally similar in shape. These results indicated, as expected, that the major difference in overall performance was that students with more schooling performed better.

Item difficulty levels for each subtest were examined for each grade in terms of average difficulty, the number of items with a less-than-chance ($p < .25$) probability of being answered correctly, and the number of moderately difficult ($p < .65$) versus moderately easy ($p > .65$) items. Average difficulties and number of chance items are presented in Table 1. The increase in difficulty levels across grades is gradual and consistent. There are a limited number of less-than-chance probability items for

eleventh and twelfth graders, with numbers increasing for tenth graders, and again for ninth graders. The shift in number of items from moderately difficult to moderately easy is also a gradual and consistent increase across grades. Even for ninth graders, the numbers are not large enough to indicate that scores are heavily influenced by guessing thereby reducing reliability.

Table 1

Item Difficulty Levels

	Grade			
	9th	10th	11th	12th
Subtest	(n=4,238)	(n=3,878)	(n=3,263)	(n=2,682)
GS	.46 (3)	.52 (3)	.56 (2)	.58 (2)
AR	.44 (3)	.50 (1)	.55 (1)	.58 (0)
WK	.56 (4)	.62 (0)	.67 (0)	.71 (0)
PC	.50 (1)	.55 (0)	.60 (0)	.63 (0)
AS	.37 (6)	.42 (2)	.46 (1)	.49 (1)
MK	.40 (4)	.46 (4)	.51 (1)	.53 (0)
MC	.40 (1)	.44 (0)	.47 (0)	.49 (0)
EI	.36 (4)	.39 (4)	.43 (2)	.45 (1)

Note. Entries are means of proportion answering each item correctly. Numbers in parentheses are the number of items with $p < .25$.

Reliabilities, as assessed by coefficient alpha, were computed for the four grades for the eight power subtests (see Table 2). Reliabilities increased at each grade for each subtest with scores being the least reliable for ninth graders and most reliable for twelfth graders. In general, the biggest differences occurred between ninth and tenth graders, with reliabilities of tenth graders being very similar to those of eleventh and twelfth graders. Except for EI, reliabilities fell below .70 only twice--ninth grade reliabilities for AS and MC were .69. Reliabilities were generally high (above .80) for subtests measuring verbal and quantitative skills. As expected, they were a little lower for PC, which has only 15 items. The technical subtests (AS, MC, and EI) had slightly lower reliabilities, especially for the lower grades. These lower reliabilities can be traced to female test-takers. In the most extreme case, EI reliabilities for males ranged from .60 for ninth graders to .73 for twelfth graders; for females they ranged from .30 to .41. Reliabilities were lower on technical subtests for females, but this was true for females of all grades.

Table 2

Subtest Reliabilities

	Grade			
	9th	10th	11th	12th
Subtest	($\underline{n}$=4,238)	($\underline{n}$=3,878)	($\underline{n}$=3,263)	($\underline{n}$=2,682)
GS	.78	.80	.81	.83
AR	.80	.84	.87	.88
WK	.85	.88	.89	.90
PC	.72	.76	.78	.79
AS	.69	.76	.80	.83
MK	.74	.80	.85	.87
MC	.69	.76	.78	.80
EI	.50	.60	.63	.66

The final analysis was a count of the number of items on each subtest that were omitted (not attempted) by greater than five percent of the students in a grade. There were few items reaching this criterion overall, with no items falling into this category for at least half of the subtests for each grade. Ninth graders had the most items not attempted, but the numbers again were low enough to indicate minimal impact on subtest scores.

Conclusions

The analyses reported here converge on the idea that ASVAB scores for ninth and tenth graders differ from scores for eleventh and twelfth graders, but that the test is appropriate for all four grades. Beginning with the comparison of mean subtest scores and score distributions, it is clear that there is a consistent yet gradual increase in performance at later grades. The similarity of the score distributions across grades confirms that there is nothing unusual about the performance of the ninth or tenth graders--the subtests allow a full distribution of ability to be realized at these grades. Item difficulty levels show the same increase across grades (with items being easier for those in higher grades), but the steady increase again suggests a difference of amount, not kind. Although items with less-than-chance probabilities are noticeable for ninth graders but not for the others, a minority of items are affected. Many of these items are accounted for by the items which were not attempted (omits), which again is very small. Finally, subtest reliabilities reveal the same increasing trend as in the other analyses. The lowest reliabilities are for the technical subtests (AS, MC, and EI) for ninth graders, and analyses by sex showed low reliabilities for females of <u>all</u> grades on these subtests. There is therefore no reason to single out ninth graders.

Although the analyses here were done at the subtest level, scores are reported to students in terms of composites. Composites are more stable estimates of performance than subtests, and their reliabilities tend to be higher. The examination of subtest scores is useful in that is makes it clear that information from the technical subtests should be used with caution, especially for females and ninth graders. Low scores on these subtests should not be used to direct students away from these areas, since many of these students are those who have not been exposed to the information tested in these subtests. The best use of the ASVAB for ninth and tenth graders is for relative information about how students in the same grade perform.

In sum, ASVAB scores for ninth, tenth, eleventh, and twelfth graders seem to provide similar information. This research suggests that the use of the ASVAB to obtain quantitative and verbal information is warranted for all grades. It should be noted that only one issue concerning the development of norms for ninth and tenth graders is examined here. Other factors must be considered before normative information is made available for ninth and tenth graders.

Acknowledgement

I am grateful to Dr. Malcolm James Ree for his technical guidance and oversight of this work.

Reference

Maier, M. H., & Sims, W. H. (1982). Constructing an ASVAB score scale in the 1980 reference population (Memorandum 82-3188). Alexandria, VA: Center for Naval Analyses.

Biographical Screening of Military Applicants:
Experimental Assessment of Item Response Distortion

Thomas T. Trent
David C. Atwater
Norman M. Abrahams

Navy Personnel Research and Development Center[1]

Abstract

The extent of faking on biographical questionnaire (BQ) items and methods to minimize faking were investigated. Subjects completed a BQ under experimentally induced sets designed to elicit faked and honest responding. The respondents demonstrated relatively moderate faking in a simulated applicant environment. Faking was reduced by using: 1) increasingly strong warning statements cautioning against faking, 2) verifiable items, and 3) an empirically derived scoring key.

Introduction

Background

Considerable interest has developed for the use of biographical data (biodata) in screening for military service (Eitelberg, Laurence, Waters & Perelman, 1984). One area of interest has focused on identifying applicants with the highest propensity for adapting to military life as measured by successfully completing obligated service. A problem in utilizing biodata for this purpose concerns the potential for inaccurate or exaggerated self-report information (Walker, 1985). Distortions of this type can increase screening scores and invalidate the screening instrument. Haymaker and Erwin (1980), for example, have reported that "fake good" distortion improved scores on the Army's Military Applicant Profile (MAP).

The purpose of this study was to gather additional information concerning the extent and nature of biodata faking in a military setting, and to examine methods of limiting such faking.

Approach

Instrument and Tenure Key

A biodata questionnaire was constructed from 50 experimental items that were considered to be susceptible to faking. The items addressed personal, school, and work experiences in a multiple choice item format. Forty-seven of

---

[1]The opinions expressed in this paper are those of the authors, are not official, and do not necessarily reflect the views of the Navy Department.

the items had response alternatives which had been empirically weighted. These weights (-1, 0, +1) were derived in a previous study (Atwater & Abrahams, 1983) which correlated item alternatives with a military tenure criterion (i.e., completion of the first six months of military service). An individual's score on this tenure key consisted of the sum of the weights from the 47 scored items.

Subjects

The respondents were 487 male Navy recruits in the third week of basic training. The biodata items were administered during a regularly scheduled "post-enlistment verification" testing period to 16 training companies ranging in size from 26 to 72 recruits. This administration period was selected because it provided a realistic environment in which to create a simulated "applicant" response set. Data from 35 respondents were excluded from the analyses because of incomplete response or non-compliance with procedures.

Design and Procedure

The experiment employed a two-factor (3 X 2) mixed design. The between-subjects factor consisted of three levels of increasingly severe warnings concerning the potential verification of responses and the consequence of detected faking. The warnings were given by written directions in the questionnaire booklet. The warning statements, derived from a pilot study, consisted of (1) no warning, (2) a moderate warning, and (3) a severe warning. The within-subjects factor was a repeated administration of the same items under two conditions involving varying response sets -- the items were administered in a simulated "applicant" environment at time-one and again at time-two in an anonymous "honest" environment. The response sets were established by verbal and written instructions as described below.

Experimental Subjects. The recruits assembled in a formal testing room expecting routine verification of certain applicant test data. The subjects were instructed to complete a new screening questionnaire that they had "missed" during their pre-service application for enlistment. They were also told that "their scores would be used to determine their suitability for Navy service." These "applicants" (N = 351) were instructed to read the questionnaire directions, which randomly contained one of the three warning conditions. Thus, at time-one the three experimental groups were answering under an "applicant" set, and presumably would be trying to "look good."

At time-two, the true purpose of the experiment was revealed to these "applicants." Each experimental subject was then told to remove his name from his time-one answer sheet. Time-two directions for the experimental subjects were to complete the questionnaire again under an anonymous and "honest" set.

Control Subjects. An independent control group (N = 101) was randomly preselected, told the true purpose of the experiment, told not to put their names on the answer sheet, and asked to respond honestly at time-one. At time-two, the control subjects were asked to retake the questionnaire, but to fake to try to "look good."

Results

Table 1 reveals that, irrespective of warning conditions, "applicants" scored significantly higher on the tenure key at time-one as compared to their anonymous "honest" responses at time-two ($F = 100.77$, $p < .001$). However, their scores as "applicants" were considerably lower than the scores of the "fake good" control group, suggesting that the "applicants" were somewhat restrained in their faking relative to the control group's artificially induced extreme faking. The applicants' scores were also reduced as a function of verification warning statement intensity ($F = 2.73$, $p < .034$). The severe warning group's time-one scores were nearly identical to those of the control group responding under anonymous "honest" conditions, indicating that verification warnings encourage honest self-reports.

Table 1

Tenure Key Score by Warning Statement Group and Response Set

		Response Set				
		Time-One (T1)		Time-Two (T2)		
Warning Statement Group	(N)	Mean	SD	Mean	SD	Test-Retest r
		"Applicant"		"Honest"		
None	115	19.04	8.15	15.57	10.50	.76
Moderate	117	18.03	7.53	14.52	8.76	.80
Severe	119	16.63	8.05	13.64	10.11	.78
Total	351	17.89	7.95	14.57	9.82	.78
		"Honest"		"Fake Good"		
Control	101	16.52	8.28	23.74	6.74	.28

Table 2 exhibits response set difference scores for the verification warning groups (T1 - T2) and the control group (T2 - T1). When instructed to fake responses in order to improve their honest scores, the control subjects changed (responded differently at retest) a relatively large number of item responses (mean = 24 changes). The applicants, however, made an average of only 9 changes. The large standard deviations (SD) found for all three of the difference calculations in Table 2 indicate a wide range of response distortion around relatively moderate mean faking.

Table 2 also demonstrates that the net result of response distortion on tenure key scores is proportionally small. For example, the no-warning group's response distortion resulted in a mean number of changes of 9.32 (out of 50 possible changes) and a mean total tenure key change of 7.91 (calculated as the sum of the absolute value of each item's key score change from time-one

to time-two). The mean net change, however, was only 3.46, indicating that some item changes resulted in a loss of scoring points rather than a gain.

Table 2

Response Set Difference Scores by Warning Statement Group

	Applicant Score (T1) Minus Honest Score (T2)					
			Tenure Key			
	Number Items Changed		Net Change		Absolute Change	
Warning Group	Mean	(SD)	Mean	(SD)	Mean	(SD)
None	9.32	(7.41)	3.46	(6.85)	7.91	(7.49)
Moderate	8.80	(7.10)	3.51	(5.29)	7.46	(6.39)
Severe	8.71	(7.31)	2.99	(6.35)	7.04	(7.32)
Control (T2-T1)	24.02	(11.00)	7.23	(9.10)	20.77	(10.37)

Table 3 compares net scoring key change and average item absolute change between 13 potentially verifiable items and 34 nonverifiable items. Both sets of items were faked, as reflected in significant differences between the applicant and honest responses ($p < .001$). The average item change, however, was significantly less for verifiable items ($t = 8.82$, $p < .001$).

Conclusions

Relative to the controls who were asked to fake, the majority of respondents in the "applicant" groups demonstrated restrained faking. The data indicate that the inclusion of verification warnings in the questionnaire resulted in a significant, though small, decrease in faking. Verifiable items were also found to be faked to a lesser degree than nonverifiable items.

The tenure-referenced scoring key was found to minimize the impact of reponse distortion on scores. This resulted from the empirical method of scoring key construction which produced a key that was not entirely consistent with the recruits' "rational" assessment of how weights were assigned. This apparently confounded the attempts of respondents to improve their scores.

The effect of faking on the potential predictive power of these biodata items is unknown. To the extent that biodata instruments demonstrate predictive validity within the context of actual response distortion, applicant fa'ing may have little practical impact. However, if the validity of biodata is enhanced by a reduction in response distortion, then methods of controlling faking, such as those suggested by this study, will be of value.

Table 3

Response Set Difference Scores by Verifiability of Items

		Tenure Key Change (T1 - T2)					
		Net Change				Average Item ABS Change	
Response Set	N	Mean	(SD)	t value	p<	Mean	(SD)
		Verifiable Items (13 questions)					
Applicant (T1)	351						
		0.81	(1.94)	7.85	.001	0.11	(0.15)
Honest (T2)	351						
		Nonverifiable Items (34 questions)					
Applicant (T1)	351						
		2.50	(4.81)	9.76	.001	0.18	(0.17)
Honest (T2)	351						

References

Atwater, D.C., & Abrahams, N.M. (1983). Adaptability screening: development and initial validation of the Recruiting Background Questionnaire (RBQ). (Report No. 84-11). San Diego, CA: Navy Personnel Research and Development Center.

Eitelberg, M.J., Laurence, J.H., Waters, B.K., & Perelman, L.S. (1984). Screening for service: aptitude and education criteria for military entry. Washington, DC: Office of Assistant Secretary of Defense.

Haymaker, J.C., & Erwin, F.W. (1980). Investigation of applicant responses and falsification detection procedures for the Military Applicant Profile. (Final Project Report, Work Unit No. DA644520). Alexandria, VA: Army Research Institute.

Walker, C.B. (1985). The fakability of the Army's Military Applicant Profile (MAP). Paper presented at the Human Resources Management and Organizational Behavior Association.

The Influence of College Education on Standardized Test Performance: Should Multiple Conversion Scales Be Used?

Ben Roach
University of Notre Dame

Abstract

The Air Force Officer Qualifying Test (AFOQT) is used in the selection process for candidates in recommissioning training The AFOQT is given to individuals with diverse levels of formal education; ranging from high school students to graduates of doctoral programs. Early studies suggested using multiple conversion scales to account for educational differences. This study concludes that the magnitude of in fluence on test performance is not so large as originally believed. Thus, a single conversion scale is acceptable and preferable.

Introduction

Most standardized tests of academic ability are targeted to a rather well defined population. For example, the California Achievement Test (CAT) is administered to high school freshmen; the Scholastic Aptitude Test (SAT) is designed for high school seniors and graduates applying for admission into college; and, the Graduate Record Examination (GRE) is designed for college seniors and graduates applying for admission into graduate school. However, the Air Force Officer Qualifying Test (AFOQT) is administered to individuals having a wide range of formal education, ranging from high school seniors to examinees with Master's and Doctoral degrees. While the influence of education on test performance may not be very important for users of most standardized tests of academic ability, it is of primary concern in the accurate interpretation of AFOQT scores.

The problem addressed by this study is actually two-fold. Does the attainment of post-high school formal education influence performance on the AFOQT? If so, what is the magnitude of the influence? This influence is operationally defined as a change in test performance which can be directly attributed to an increase in college education.

Background

As the Air Force began to develop into a major branch of the Armed Forces, multiple sources for obtaining commissioned officers evolved, with the AFOQT being administered to each applicant regardless of precommissioning program. Thus, mean scores on the AFOQT were used to provide an "index of quality" for comparing different groups based on source of commission. Apparent differences in mean score performance between groups were observed, but these were regarded to be a function of the different educational levels represented by each of the sources, i.e., college freshmen for Air Force Reserve Officer Training Corps (AFROTC) and college graduates for Officer Training School (OTS).

Gregg (1968) performed a study to investigate these differences. His study was based on the hypothesis that "maturation and educational experience will have an elevating effect on the score achieved by an individual on the AFOQT" (p.2). To test this hypothesis, the AFOQT was administered to a sample of 415 AFROTC cadets during their senior year. Since each cadet had been tested operationally as a freshman or sophomore, the scores from the experimental test administration (test 2 as seniors) were compared with those obtained operationally (test 1 as freshmen). The mean increase in performance for all pairs of scores was taken as a measure of the increase in performance due to increasing the educational level. One of the major conclusions of the study was that college seniors score <u>a minimum of 30 percentile points higher</u> than college freshmen on all composites of the AFOQT.

To confirm the Gregg study and to increase confidence in differences between conversion tables and qualifying scores, Tupes and Miller (1969) performed another evaluation. The purpose of their study was to investigate educational effects on AFOQT scores by use of "a research strategy markedly different" from the Gregg study. Their approach was to use another test as an independent measure of aptitude. At the time of this second investigation, the Department of Defense Officer Record Examination (DORE) was being administered to junior officers from nearly all sources of commission. The DORE was a high-level academic aptitude test which contained verbal and quantitative subtests. Thus, the content was considered essentially equivalent to the academic composites of the AFOQT, e.g., Officer Quality, Verbal, and Quantitative.

The sample for the Tupes and Miller study consisted of all examinees who were administered the DORE during Fiscal Year 1966 and whose source of commission was AFROTC or OTS. AFOQT scores were matched with DORE scores on 2,840 AFROTC commissionees and 685 OTS commissionees. Using the DORE scores as a way of equating the groups, the conclusion of this study was that, "the increase in AROQT performance as a result of about three years of formal education at the college level is from approximately 10 to 30 percentile points for the various aptitudes measured by the test" (p. <u>v</u>). Since their findings confirmed those of the Gregg study, separate percentile tables were constructed for AFROTC and OTS applicants.

In response to the need for a more complete research data base, AFHRL undertook a project to rescore a large number of AFOQT test records in order to obtain subscale and item response data. The AFOQT rescoring effort resulted initially in the development of a new data base containing scores on about 68,000 examinees who had taken a single form (Form N) of the AFOQT from 1978 through the first-half of 1980. Evaluation of this data base revealed that the mean score differences between college freshmen and college graduates was not the 10 to 30 or more percentile points expected but only about 4 percentile points on the Academic Aptitude (formerly called Officer Quality Composite, Roach, 1981). Thus, the justification for the use of multiple conversion tables and different qualifying minimum scores for each of the sources of commission was questionable.

The 1981 study differed from the previous studies on three important aspects: (a) number of test forms used, (b) experimental design, (c) and type of subjects. Number of test forms used refers to single or successive forms. The 1981 study used examinees on a single form (Form N), the 1968 study used two successive forms (AFOQT-64 and AFOQT-66), and the 1969 study used data from three successive forms (Form G, AFOQT-64, AFOQT-66). Experimental design refers to the way in which data were compared. The 1981 study used a cross-sectional analysis of a single large sample, the 1968 study was longitudinal,

and the 1969 study treated two groups as equivalent at the time of DORE testing (which assumed no group differences due to selection procedure or training difficulty). The type of subjects also differs with the three studies. The 1981 study used an applicant population (unrestricted), the 1968 study used AFROTC cadets (partially restricted), and the 1969 study used AFROTC and OTS commissionees (fully restricted).

Method

Accurate resolution of the problem requires a longitudinal study. A true experimental design using randomly selected subjects would be the preferred approach. Such an undertaking would require four to six years to complete and an excessively large amount of administrative overhead. Thus, a quasi-experimental design using archival data was used as an alternative.

After eliminating test records which had missing or inaccurate information on education level at the time of testing, a consolidated data base of about 145,000 cases was reduced to 82,845 individual test records and 4,783 paired test records. This consolidated data base contained results on these successive forms of the AFOQT (Forms L, M. and N). Percentile scores based on the Common Metric scale were used for each of the five composites of the AFOQT test battery. The Common Metric is a percentile scale on which all successive forms of the AFOQT used in the data base were equated. Means and standard deviations for the 82,845 individual test records tend to support the findings of the 1981 cross-sectional analysis. For example, individuals with 12 years of education(n = 25,889) had a mean percentile score of 54.8 on the Academic Aptitude composite. Those with 16 years of education (n = 26,063) had a mean percentile score of 56.6 percentile points.

While a difference in mean score of only 1.8 percentile points is consistent with the earlier cross-sectional analysis, the influence on test performance which can be directly attributed to an increase in college education is not adequately addressed. To do this, the first subsample (n = 2,447), composed of individuals who had paired-test records and who had increased their forman education level, was categorized into four stratified experimental groups based on the amount of change in education level between the first test and the second (one-,two-,three-, and four year increase).

The paired-test records for each individual within each experimental group provided a performance measure known as a difference score. The difference score is operationally defined as the result obtained by subtracting the pretest score from the posttest score. A mean difference score was then computed for each of the five composites of the AFOQT for each of the four experimental groups.

The second subsample, composed of individuals who had paired-test records but who had not increased their educational level (n = 2,336), was used as a control group to estimate the influence of the test-retest effect. This effect is operationally defined as the interaction of the passage of time between pretest and posttest measured in months. Simple linear regression equations were developed from the control subsample for each of the five composites of the test battery. These equations were then used to predict the test-retest effect for time periods representative of those used for each group of the experimental subsample (12, 24, 36, and 48 months). Deviations between actual mean difference scores obtained for each group of the experimental subsample and corresponding predicted mean difference scores obtained from the control subsample provide a measure which is considered to be the influence of college eduction. These deviations are considered to be a direct

indication of the influence of college education on test performance because the mean difference scores have been corrected for the estimated effects of test-retest.

Results

Education level is the key variable in this study. Individuals who had increased their formal education level between two administrations of the AFOQT were compared with those who had not increased their eduction level during a comparable passage of time. Using the reported education level at the time of the first test (pretest) and at the time of the second test (posttest), 25 sets of test-pairings are possible. Differences for most of the 25 test-pairings were minimal. However, the mean difference score increased over 11 percentile points after only one additional year of formal education for both the Pilot and the Navigator-Technical composites. Changes in score for these composites were much lower for 2, 3, and 4 years of additional education. This large increase in test performance after one year appears to be due to familiarity with the test since individuals who had no reported change in formal education had as much of an improvement in scores as those with a one year change in education level.

The mean difference scores were very low for all five groups on the Academic Aptitude composite indicating, at least initially, that this composite is not affected by changing educational levels. However, the Academic Aptitude composite is obtained by combining raw scores from the Verbal and Quantitative composites. The mean difference scores on the Verbal composite increased in direct proportion to an increase in years of formal education. These increases are negated by the inversely related changes on the Quantitative composite. Thus, while additional years of formal education appears to increase performance on the Verbal composite, performance on the Quantitative composite decreases.

To determine if these changes are actually due to the influence of formal education or simply the passage of time, regression equations were computed from the group of subjects reporting no change in formal education level. Subjects in the control group were arrayed on the basis of the number of months elapsed between pretest and posttest. The mean difference score was computed for all subjects within each element of the array up to 60 months. A separate regression equation was computed for each composite of the AFOQT. These equations were then used to provide a predicted mean difference score for 12, 24, 36, and 48 months. The four predicted scores on each composite correspond to the 1, 2, 3, and 4 years of additional education used to create the four experimental groups. The passage of time results in a decrease in performance on all five composites. The rate of decrease was smallest for the Verbal composite (-.03 percentile points per month) and largest for the Quantitative (-.32 percentile points per month). The rate of decrease was almost identical for the Pilot, Navigator-Technical, and Academic Aptitude composites (-.14, -.15, and -.16, respectively).

The final step in the procedure was to correct the actual mean difference scores for test-retest effects. This was accomplished by subtracting the predicted mean difference scores from the actual mean difference scores. This provided a mean difference score which indicates the influence of formal education on AFOQT test performance. There is a slight upward trend for both the Academic Aptitude and Verbal composites. The total change after four years of formal education amounts to about a one percentile point increase per year of additional education, hardly enough to justify multiple conversion tables.

There are no apparent trends with any of the three remaining composites. Table 1 shows the corrected mean difference scores for each of the five composites.

Table 1
Corrected Mean Difference Scores

	AFOQT Composite				
Change in Technical	Aptitude	Navigator- Verbal	Academic Quantitative	Educational Level	Pilot
1 year	0.03	1.52	-1.86	-0.29	3.77
2 years	-8.70	-2.01	-0.53	1.18	-1.75
3 years	-11.92	0.13	0.11	2.52	-1.63
4 years	-6.88	3.30	4.69	5.45	3.75

Discussion

The findings of this study do not support the results of studies performed 15 years ago. Individual performance on two of the five composites appears to be influenced by four years of formal education, but the increases observed are much too small to justify the use of multiple conversion scales.

As as result of the findings reported here, the current operational version of the AFOQT uses a single conversion scale for each of the five composites comprising the test battery. Additionally, test records obtained since 1972 have been rescored using revised conversion tables. Thus, all current applictions in the officer selection procedure, regardless of when they may have taken the test or what their education level was at the time of testing. It should also be noted that revisions in Air Force regulations have helped to ensure that test scores are recent and, therefore, applicant groups are more homogeneous with respect to education levels at the time of evaluation.

References

Gregg, G. (1968). The effect of maturation and educational experience on Air Force Officer Qualifying Test scores. (Technical Report AFHRL-TR-68-107). Lackland AFB, TX: Air Force Human Resources Laboratory, Personnel Research Division. (DTIC No. AD-687 089)

Roach, B. W. (1981). AFOQT - One year later. Paper presented to AU/AFROTC Joint Conference on Officer Selection, Air University, Maxwell AFB, AL.

Tupes, E. C. & Miller, R. E. (1969). Equivalence of AFOQT scores for different educational levels. (Technical Report AFHRL-TR-69-19). Lackland AFB, TX: Air Force Human Resources Laboratory, Personnel Research Division (DTIC No. AD-703 727)

Selected Characteristics of High School Students Who Indicated Their Post High School Plans Were Influenced A Great Deal by Military Recruiters

John W. Skehan
University of Maine at Orono

Abstract

An analysis was made of selected responses of selected high school seniors who participated in the longitudinal study entitled High School and Beyond (National Center for Education Statistics, 1983). A crosstab computer analysis was made of selected responses of 548 students who reported that their after high school plans were influenced "a great deal" by military recruiters. The purpose of the analysis was to determine the personal, familial and educational characteristics of that group of students - students who would appear to be prime candidates for military recruiters. Findings indicate that these students appear to be less academically oriented in high school and have lower educational aspirations than students whose after high school plans were "not at all" influenced by military recruiters. The parents of these 548 students have less education, lower incomes and less skilled occupations than the parents of students who were "not at all" influenced by military recruiters.

Introduction

Recently education researchers at the University of Maine at Orono gathered and analyzed a large quantity of data to determine the aspiration levels of Maine's rural high school students. In addition, data collected from a longitudinal study entitled High School and Beyond (National Center for Education Statistics, 1983) were analyzed to determine if aspiration levels of rural students nationwide differ from those of urban and suburban settings. This analysis was undertaken to determine if Maine's documented problem of low aspirations of its youth might be related to a larger, national phenomenon (Cobb, McIntire and Pratt, 1985). During a review of data from High School And Beyond (HSB), this author became aware that the HSB study contained data which might be of assistance to military recruiters. This paper reports on selected data from the HSB study.

Method

The High School and Beyond (HSB) data were collected by the National Center for Education Statistics (NCES) "...to study longitudinally the educational, vocational, and personal develop-

ment of high school students and the personal, familial, social institutional, and cultural factors that may affect that development" (NCES, 1983). In the 1980 sample of the HSB data base, students were selected through a two-stage stratified probability sample with schools as the first stage units and students within schools as the second stage units of the sampling procedure. Strata used in the school sampling included nine U.S. census regions; size of enrollment; racial composition; urban, suburban or rural settings, and public, private or parochial focus. A total of 1,015 schools were selected for the sample from a total of 24,725 schools. Within each school, 36 seniors and 36 sophomores were randomly selected to participate in the HSB data collection. In schools with fewer than 36 seniors or 36 sophomores, all eligible students were included in the sample. This paper reports on an analysis of selected data from 10,829 seniors.

One of the HSB questions asked seniors "How much has each of the following persons influenced your plans after high school?" The persons listed included father, mother, teachers, guidance counselor, friends or relatives about the student's own age, military recruiters and college recruiters. Possible responses included "not at all", "somewhat" and "a great deal". Five hundred forty-eight (5.1%) respondents reported that military recruiters had a "great deal" of influence on their plans after high school, 1,246 (11.5%) said that military recruiters had "somewhat" of an influence and 9,035 (83.4%) indicated that military recruiters had influenced their plans "not at all". This paper presents an analysis of selected responses of the 548 seniors who reported that their plans after high school were influenced "a great deal" by military recruiters. The author utilized a computer crosstab analysis of certain selected variables and student responses to the question indicating how much influence military recruiters had on their after high school plans. The purpose of this analysis was to determine the personal, family and educational characteristics of that group of students - students who would appear to be prime candidates for military recruiters.

Findings

Family income level and the educational level of parents appear to be factors which relate inversely to the influence military recruiters have on the after high school plans of students. In general, the higher the family income and/or the higher the level of parental education, the less likely it is that a student's after high school plans will be influenced "a great deal" by military recruiters. With respect to family income, students from families with incomes below $16,000 are three to four times more likely to be influenced "a great deal" by military recruiters than students from families with incomes of $25,000 or more (see Table 1). If neither parent, step-parent or guardian has a college degree, a student is nine times more likely to be influenced "a great deal" by military recruiters than if either parent, step-parent or guardian has a college degree (see Table 2).

Table 1

Family Income of Students Whose After High School Plans Were Influenced A Great Deal by Military Recruiters

	%
$ 6,999 or less	19.6
7,000 - 11,999	19.2
12,000 - 15,999	20.0
16,000 - 19,999	16.3
20,000 - 24,999	13.6
25,000 - 37,999	6.6
38,000 or more	4.7

Table 2

Parental Education Level of Students Whose After High School Plans Were Influenced A Great Deal by Military Recruiters

	% Mother	% Father
Less than bachelors degree	93.8	92.9
Bachelors degree or greater	6.2	7.1

The extent of parental influence, especially that of the mother, on the student's after high school plans appears related to the degree of influence that military recruiters have on those plans (see Table 3). A majority of students who are influenced "a great deal" by military recruiters also have parents, especially mothers, who support the students' after high school plans.

Table 3

Influence of Other Persons on the After High School Plans of Students Whose Plans Were Influenced A Great Deal by Military Recruiters

	% Not at All	% Somewhat	% Great Deal
Father	21.6	35.1	43.3
Mother	9.4	32.4	58.2
Teachers	32.1	37.1	30.8
Counselor	32.0	37.7	30.3
Friends/Relatives	16.4	46.3	37.3

Parental occupations of students whose after high school plans are influenced "a great deal" by military recruiters appear related to the educational levels of the parents. Relatively few of these students have parents who are involved in professional occupations, a not-too-surprising finding, since relatively few parents are educated to the bachelors degree level or greater. The three most frequently reported occupations for fathers are craftsman (14.6%), operative (13.1%) and laborer (12.8%), while the three most frequently reported occupations of mothers are service (14.8%), homemaker or housewife only (11.5%) and clerical (11.3%)(see Table 4). Approximately five times as many of these students do not live with their fathers (15.0%) compared to those who do not live with their mother (3.1%). Conversely, nearly twice as many of these students do not know their mothers' occupations (14.6%) compared to those who do not know their fathers' occupations (7.5%).

Table 4

Most Recent Occupations of Parents of Students Whose After High School Plans Were Influenced A Great Deal by Military Recruiters

Occupation	% Father	% Mother
Do not live with parent/guardian	15.0	3.1
Clerical(secretary, teller, mailman)	1.1	11.3
Craftsman	14.6	2.0
Farmer, farm manager	4.7	.9
Homemaker or housewife only	.5	11.5
Laborer	12.8	2.4
Manager, administrator	5.8	3.8
Military	3.6	.2
Operative	13.1	6.4
Professional (accountant, etc.)	2.6	6.9
Professional (clergy, dentist, lawyer)	1.3	.5
Proprietor or owner	2.6	1.3
Protective service	1.6	
Sales	1.3	3.3
School teacher	.7	4.0
Service	1.5	14.8
Technical	1.6	.9
Never worked	.9	6.4
Don't know	7.5	14.6
No answer, multiple punched answers, etc.	7.2	5.7

Nearly half (43.4%) of the students who are influenced "a great deal" by military recruiters are enrolled in the general course in high school, one-third (32.9%) are enrolled in various vocational programs, and slightly less than one-quarter (23.6%) are enrolled in academic or college preparatory courses. Slightly more than one-third (36.5%) of these students reported that they "got mostly A's and B's" in mathematics courses, and slightly less than one-third (31.2%) reported that they received such grades in English.

The educational aspirations of students who are influenced "a great deal" by military recruiters are much lower than those of students who report being "not at all" influenced by military recruiters. Nearly one-half (43.6%) of the former group indicate that graduation from high school is the lowest level of education with which they would be satisfied, while only about one-quarter (26.4%) of the latter group would be satisfied with that level of education. Less that one-fifth (15.9%) of the former group report that the lowest level of education which would satisfy them would be to complete college and/or an advanced degree. On the other hand, more than one-third (33.6%) of the latter group would be satisfied with nothing less than completing college or an advanced degree.

Conclusions

On the basis of the findings reported herein, it appears that students whose after high school plans are influenced "a great deal" by military recruiters tend to be those students who are less academically oriented in high school and whose educational aspirations are much lower than students whose after high school plans are "not at all" influenced by military recruiters. The parents of students influenced "a great deal" by military recruiters tend to have less education, lower incomes and less skilled occupations than the parents of students who are "not at all" influenced by military recruiters.

References

Cobb, R.A., McIntire, W.G, and Pratt, P.A. Where are the Dreamers? Explorations: A Journal of Research at the University of Maine at Orono, 1985, 1, 6-11.

National Center for Education Statistics (NCES). Contractor Report: Data File User's Manual for the High School and Beyond 1980 Sophomore/Senior Cohort First Followup (1982). Washington, D.C.: U.S. Department of Education, 1983.

Common Metric: A Single Scale for Comparing Different Applicant Groups

Ben Roach
University of Notre Dame

Deborah L. Rogers
Air Force Institute of Technology

Abstract

Development and standardization of a single conversion scale is reported. This scale, known as the Common Metric, is used to equate successive forms of the Air Force Officer Qualifying Test (AFOQT). The common metric allows individual and group performance to be compared across past, present, and future AFOQT forms.

Introduction

This paper presents the rationale, development, and standardization of the Common Metric, a single scale used to report observed scores on successive forms of the Air Force Officer Qualifying Test (AFOQT). The structure of the report is two part: first, a discussion of the rationale of the Common Metric and development of the Form N reference group used as the normative base; and second, a discussion of the development of Common Metric scales for two previous versions of the test battery, AFOQT Forms L and M. Since this paper is a report on the development of a Common Metric, and not a report of a formal *experiment, a results section* is not included.

Background

A standardized test for selecting officers into the Air Force has been in use since World War II. The first selection instrument, the Aircrew Classification Battery (ACB), used the mobilization population as a norming base. The end of the war resulted in a discontinuance of the ACB and the mobilization population. When officer selection testing resumed in 1949, the norming base was composed of West Point cadets. Additionally, the Air Force Officer Qualifying Test (AFOQT) was developed to replace the ACB.

By 1955, the AFOQT was normed using cadets at the newly initiated Air Force Academy (AFA) as a reference group. The AFA cadets were used until 1960 when the requirement for the AFOQT as a selection test was eliminated in favor of the College Entrance Examination Board (CEEB). In anticipation of the loss of the AFA applicant population as a reference group, a new normative base was obtained by administering Form G of the AFOQT and the Project TALENT test battery simultaneously to more than 5,000 applicants for the AFA class entering in 1960. A subsequent indirect method of using TALENT composites and basic airmen became the accepted procedure for standardizing successive forms of the AFOQT.

Over the next twenty years, numerous adjustments were made which resulted in multiple conversion tables for each successive form of the AFOQT. By 1980, multiple changes to norming and standardizing the test battery had created a situation in which AFOQT scores were no longer comparable across groups of applicants. Since there was no time limit on how long AFOQT scores would be considered valid, an individual could apply for a precommissioning program

with scores that were up to 10 years old. Therefore,an applicant's percentile score could be obtained from one or more of the 54 different conversion tables accompanying four successive forms of the test battery.

For example, suppose three applicants applied for OTS at the same time and all three have the same "true" score on the Officer Quality (now called Academic Aptitude) composite, but the sex, education level, and form of the test differed. Assume applicant A is a male who took Form L of the AFOQT in 1982 as a college freshman, applicant B is a female who took Form M in 1976 as a college senior, and applicant C is a male who took Form N in 1979 as a college graduate. The Officer Quality Percentile scores obtained are 40 for applicant A, 25 for applicant B, and 50 for applicant C. This means that applicant A performed better than 40% of all male college freshmen taking Form L, applicant B performed better than 25% of all female college seniors taking Form M, and applicant C performed better than 50% of all college graduates taking Form N of the AFOQT. While it is possible to determine the relative standing of applicants within their respective peer groups, it is not possible to accurately rank-order these applicants if all are combined into a single applicant group. While all three applicants were of equivalent "true" ability, applicant C, under the old system, would have been ranked higher than applicant A, and applicant A ranked higher than applicant B.

In other words, as long as the selection process occurred in an ideal setting with a homogenous group of applicants, the scores would be meaningful. But in the real selection environment, homogeneous applicant groups do not normally exist. Therefore, the purpose of this study was to establish a single scale that can be used to interpret the scores of all applicants taking recent, current, and future forms of the AFOQT. This scale, referred to as the Common Metric, allows for an accurate comparison of individual and group scores.

Approach

Prerequisites

Standardization of the AFOQT must provide uniformity in the meaning of test scores without regard to the time or place of testing. This uniformity requires that a single reference group be the basis of all conversion tables. Therefore, the development of a common metric was a two-stage process. The first stage was the development of a suitable reference group, adequately described, to serve as the norming anchor for an extended period of time. The second stage was the development of the single scale which is used to standardize AFOQT test scores.

An appropriate reference (or anchor) group should satisfy five basic prerequisites. First, the group must be adequately defined to serve as a standard for reference, e.g., the World War II mobilization population, the applicants for the AFA class of 1964, etc. Second, the group should be relevant to the intended purposes of the selection process. Third, sampling error should be minimized. The larger the sample, the more likely this requirement will be satisfied. Fourth, the group should be representative of the intended testing population. If the normative sample is biased due to disproportionate representation of subjects from particular demographic categories (e.g., sex or race groups), potentially serious error may be introduced. Finally, the influence of motivation should be controlled. This prerequisite is dependent upon the source of subjects used to develop the normative base. Often the environment of an experimental testing situation

does not encourage the individuals to perform at their full potential. For example, an examinee who has already been accepted into a precommissioning program may be less motivated than one applying for selection.

Subjects

In 1980 an effort began at the Air Force Human Resources Laboratory (AFHRL) to implement a centralized AFOQT scoring system which provided electronic scanning of answer sheets and computerized scoring for examiners who tested at sites other than those used by the Air Force Reserve Officer Training Corps (AFROTC). Since AFOQT test records had been sent to AFHRL for archival storage, it was possible to build a very large data base containing test records on all individuals given the AFOQT at non-ROTC test sites since January 1975. AFROTC had implemented a similar centralized scoring system several years earlier; thus, a data base containing the AFOQT testing records from the ROTC detachments since January 1972 was also available to AFHRL. By combining the two sets of data it was possible to build a consolidated data base which included the testing records on about 145,000 individuals who took recent forms of the AFOQT in an operational environment (i.e., actual applicants).

The consolidated AFOQT data base appeared to be an ideal source for a new normative base. The group was certainly relevant to the intended purposes of the selection process. Since it contained essentially the entire testing population, the sampling error would be minimized. The group would be unquestionably representative, and the influence of motivation was adequately controlled since the test records were obtained from operational testing environments. In an operational environment, examinees are voluntary applicants for a precommissioning or specialized training program. Thus, the examinee in an operational environment will probably be more highly motivated than one tested in a research environment, and the scores will be a more realistic measure of true ability.

The only significant limitation to use of the consolidated AFOQT data base was that it consisted of scores of examinees who had taken only one of three successive, but not equated, forms of the AFOQT. To illustrate the lack of equivalence between successive forms, three samples were obtained from the data base; each of the three samples consisted of individuals who had taken only one form of the AFOQT. These samples are referred to in this study as "operational" samples to prevent confusion with the reference group used to anchor the Common Metric scales.

The three operational samples were limited to individuals who were tested at non-ROTC test sites (i.e., civilians tested at Armed Forces Examining and Entrance Stations [AFEES] and active duty military members tested at Air Force bases). The Form L operational sample consisted of 2,007 examinees tested from January 1975 through April 1975. The Form M operational sample contained 9,673 examinees who were also tested from January 1975 through April 1975. Form M of the AFOQT was implemented in January 1975 (4 months before use of Form L was discontinued). Thus, these two samples are representative of the transition period that occurs when a new form of the AFOQT is implemented on a world-wide basis. The third operational sample contained 22,101 examinees administered Form N of the AFOQT from March 1980 through December 1980. By analyzing the three operational samples it was possible to determine which of the three successive forms of the test should be used as a reference.

Common Metric Reference Group

Form L of the AFOQT was administered from 1972 to 1975, but test records were available on few individuals who tested on Form L at non-ROTC test sites prior to January 1975. Test records were available on a large proportion of those who took Form M from 1975 to 1978. But this period in time was atypical for Air Force officer selection for two reasons. First, the end of the Vietnam era had produced a surplus of officers on active duty. As a result of this surplus, the OTS precommissioning program was all but closed. Second, the end of the draft and the prevasive anti-military sentiment of the mid-70's was suspected of having an adverse influence on the characteristics of the AFROTC applicant population.

Test records were available on nearly everyone who took Form N since its implementation in April 1978. The surplus of officers from the Vietnam era had been greatly reduced and OTS had increased its output. The anti-military influence was fading from the campuses, and the nation was in a period of relative economic stability. Therefore, the group of applicants taking Form N of the AFOQT at operational test sites between April 1978 and September 1980 is defined as the anchor or reference group for the new normative base. The Common Metric Reference Group serves as the anchor to which previous forms (L and M) and subsequent Forms (O, P, etc.) of the AFOQT can be linked. Use of this anchor group provides the means for comparing individuals and groups who have taken different forms of the test.

Method

After determining that examinees taking Form N would provide the appropriate sample for use as a reference group, all records from the sample were rescored to assure accuracy and to obtain item response data. Cumulative frequency distributions were then produced for each of the five AFOQT composites on Form N. The cumulative frequency distributions served as the basis for developing Common Metric conversion tables for Form N.

The Common Metric score is defined as the percentile value for the examinee on the particular composite of the AFOQT if that examinee had been in the reference group. Thus, a percentile value indicates the relative standing of an examinee when compared against a reference or norm.

Once the Common Metric conversion tables were developed for Form N, it was possible to develop Common Metric conversion tables for Form M by using test items appearing on both forms as a direct link between the two forms. The raw score distribution of Form N was equated to the raw score distribution of N(M) common items (test items on Form N derived from Form M) using the equipercentile equating technique described by Angoff (1971). Equating was performed for each composite of the test battery.

Next, the raw score distribution of Form M was equated to the raw score distribution of Form M(N) common items (test items from Form M that were used on Form N) using the equipercentile equating technique. Since the set of N(M) common items was identical to the set of M(N) common items, the raw score distributions of each could be directly linked together.

Thus, Common Metric conversion tables were constructed for each composite of Form M by equating the common items between successive forms. The logic of this process can be summarized as follows:

(1) Form N composite raw score distribution available on the Common Metric (the reference group).

(2) Form N composite raw score distribution equated to N(M) comm... raw score distribution using equipercentile technique.

(3) N(M) common items directly matched to M(N) common items.

(4) M(N) common items raw score distribution equated to Form M composite raw score distribution using equipercentile technique.

The Form M composite is now on the Common Metric. This procedure was used for each of the five composites of the AFOQT test battery resulting in five Common Metric tables for Form M.

A different logic was used in the development of Common Metric conversion tables for Form L of the AFOQT. Since there were only 15 test items common to both Forms L and M and only 9 test items common to both Forms L and N, common items could not be used to equate Form L to the Common Metric scales. However, there was a four-month overlap between the implementation of Form M and the discontinuance of Form L.

This means that two successive forms of the AFOQT were administered to the same applicant population. Since Forms L and M had the same number of total items on each composite, it was possible to equate the raw score distributions of Form L to the raw score distribution of Form M using the equipercentile equating technique.

Conclusion

The Common Metric Reference Group is composed of 58,938 subjects who were actual AFOQT examinees. This group is further defined as 82.2% males, 77.3% whites, and 35.6% college graduates. Within the reference group, 44.7% were examinees at AFROTC detachments, 36.7% were civilians tested at AFEES, and 18.6% were active duty military members tested at Air Force bases.

The development of a large reference group has been used to resolve several problems which evolved over years of AFOQT development and usage. The Common Metric now provides a means of comparing individual applicants or groups of applicants who have taken the AFOQT Forms L through N, as well as the most recent version, Form O. The Common Metric can be used on future forms of the AFOQT as they become operational.

The development of the Common Metric has made it possible to determine the relative standing of each applicant not only in the applicant's peer group, but also against a much larger applicant population. For example, an individual taking Form L of the AFOQT and scoring a 40 has essentially the same ability as an individual who takes Form N of the AFOQT and also scores 40. Thus, the Common Metric is a single scale that can be used to equate the test results of applicants who have taken past, present, and future forms of the AFOQT.

References

Angoff, W. H. (1971) Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), Educational Management (2nd ed.). Washington: American Council on Education.

Creative Leadership: Putting Words to Music

Patrick D. McDonough
Kellogg Foundation

William E. Rosenbach
Gettysburg College

Robert L. Taylor
University of Louisville

Abstract

This interactive seminar is designed for participation. The objective is to explore the theoretical notions of creativity, linking them to the leadership situation. An essential ingredient of effective leadership is vision. This session examines the variables of creative leadership through a set of unique experiential exercises.

Introduction

An inescapable aspect of leadership is insight or vision. Successful leaders are often trying to generate (or stimulate others to generate) original ideas and develop new ways of doing things (McCall, 1979). These are not necessarily better, just different. While we recognize the leader's role in articulating the future to constituents, the real challenge is generating creative alternatives.

Bennis (1982) found in a study of 90 CEOs that all possessed the capacity to create and communicate a compelling vision of a desired state of affairs, a paradigm that induces the commitment and clarity to the vision. A unique way in which we can stimulate creativity in leadership is through means that are non-traditional--artistic rather than scientific.

Method

The Kirton Adaptation-Innovation Inventory (KAI) is an instrument that can be used to examine creativity in leadership (Kirton, 1980). This theory is concerned with differences in thinking styles employed by individuals for creative problem-solving and decision-making. This concept is particularly relevant for leaders and managers who frequently must focus on

the interaction between people and their changing environments. The scope on the KAI places an individual on a continuum ranging from highly adaptive to highly innovative. The range of responses is relatively fixed and stable, and in the general population approaches a normal distribution.

Adaptors focus their efforts for change on improving and doing better. They arrive at ideas based on stretching existing definitions of the problem and likely solutions, look at them in detail, and proceed within established boundaries or norms. This style of thinking is said to describe "managers."

Innovators, on the other hand, are more likely to reconstruct the problem in their pursuit of change. They will separate information from its enveloping accepted thought, paradigms, and customary viewpoints, and propose unexpected and probably less acceptable solutions. They are less concerned with "doing things better" than with "doing things differently."

In terms of the adaptation-innovation continuum, insight is related to innovation. A method of understanding insight-innovation is through the documentary film of Ravel's "Bolero." Briefly, this film reveals the preparation and rehearsal by the Los Angeles Philharmonic Orchestra. How leadership comes about in a large, conventional organization is portrayed in a non-conventional way--how it works in a symphony. At the same time, one can see that the basic management principles are also necessary for the organization to achieve its goal.

Summary

The discussion focuses on variables about which one can learn the leadership/followership trade-offs, team building, the leitmotif as "critical issue," gestalt in problem-solving, simulation, and imaging. Understanding the insight-innovation continuum helps the participant to appreciate the value structure of leadership relative to creativity.

Obviously, there are no set patterns of response. The objective of this seminar is to highlight the notion of insight in the leadership dimension through a stimulating creative participant experience. This model may well prove useful for teaching others about the creative element of leadership.

References

1. Bennis, Warren, The art form of leadership. Training and Development Journal, April 1982, 44-46.

2. Kirton, M. J., Adaptors and innovators in organizations. Human Relations (3) 1980, 213-224.

3. McCall, Morgan W., Jr., Conjecturing about creative leaders. The Journal of Creative Behavior (14:4) 1979, 225-234.

A COMPUTER-BASED COCKPIT DESIGN AND EVALUATION SYSTEM

Kim M. Mazur, LT USAF and Larry C. Butterbaugh
Crew Systems Development Branch (AFWAL/FIGR)
Air Force Wright Aeronautical Laboratories

ABSTRACT

The Crew Systems Development Branch, of the US Air Force Wright Aeronautical Laboratories' Flight Dynamics Lab, has implemented a system of computer based human factors tools to assist crew system engineers in the design and evaluation of candidate cockpit layouts. The system employs computer graphic, analytic and data base technologies to support crewstation layout, reach analysis, function allocation and task analysis, crew workload analysis, and advanced display design. The tools are all accessed through a single, user-computer interface; a series of on screen, menu-based prompts. This eliminates many of the potential disadvantages normally associated with such an approach, such as diverse user backgrounds and varying familiarity with computers. Also, the tools have been implemented in a modular architecture to permit easy upgrade, replacement, or addition of any tool. The result is a powerful system of tools with many desirable attributes including user-friendliness and software portability.

INTRODUCTION

All man-machine interfaces (MMI) need to be designed and evaluated using a structured and comprehensive process. This requirement is even greater for complex systems, such as aircraft cockpits or other weapon system crewstations. In order to achieve the most effective design for these workstations, numerous human factors issues, such as cockpit geometry, panel layout, crew workload, display format, and crew performance, must be considered during the MMI design process.

The use of 'traditional' human factors design and evaluation tools, such as mockup evaluations, can be effective, however, they generally require large amounts of manpower and time. Also, they typically produce qualitative results which must be further validated in manned simulations.

Computer-Aided Design and Evaluation Techniques (CADETs) benefit the MMI design process by requiring fewer resources (manpower and time), producing more objective, quantitative results and allowing consideration of more design issues.

CADET Program Goals

The CADET Program has three primary goals:

1) identify specific computer-based tools for human factors design and evaluation;

2) acquire selected software and implement them on a single (if possible) host computer;
3) develop a user-interface to permit easy and standardized ac- access to the tools.

As reported previously (Ref. 1), a CADET study contract revealed that numerous computer-based human factors design and evaluation tools existed. It also identified which tools would best merge with the AFWAL/FIGR MMI design philosophy and structure.

The selection of tools was based on two factors. First, tools developed with government funds, or otherwise available without cost, were considered, because of the large software development cost savings. Second, the host computer and operating system of the tools were considered, because this also would result in a cost savings if all tools ran on the same computer.

An efficient user-interface is desired for two reasons. First, it provides a standardized interface with all the tools, minimizing the need for user familiarization with several interface structures and control languages. Second, it minimizes the need to acquire and maintain numerous pieces of hardware, by allowing accessability of any tool from a centralized design terminal.

THE CADET SYSTEM

The CADET System consists of a workstation, a host computer, design and evaluation tools, data bases, file utilities and a user-interface module. The CADET System has a modular structure which allows the implementation, use, modification and replacement of specific tools without affecting other system components.

CADET Workstation/Host Computers

The CADET workstation includes an Envision 230 advanced color graphics terminal, an Envision 420 color dot matrix printer, and a digitizing tablet. The primary host computer for the CADET software is a Digital Equipment Corporation VAX 11/780 computer with a VAX/VMS 4.3 operating system. One software package, however, due to its size, is hosted on a Control Data Corporation Cyber 750 computer with a NOS operating system. The Cyber 750 is connected to the VAX 11/780 by an intelligent switch.

CADET Software

The CADET software consists of the user-interface module, design and evaluation tools, supporting data bases, the data base manager, and VAX file utilities. The user-interface module was designed in-house to provide a standardized method to access the tools from a centralized workstation. The design and evaluation tools include: a reach analysis tool, a workload analysis tool, a system analysis tool, and a display format design tool. The data base manager, currently under development, will provide a method to share the data bases common to the design and evaluation tools. The VAX file utilities provide a way for the user to access data files.

The User-Interface Module. The user-interface module (UIM) is a simple, user-friendly interface between the user and the CADET tools and utilities (Ref. 2). The UIM is hosted on the VAX 11/780 and consists of digital control language (DCL) procedures. The UIM provides CADET users, few who are computer experts, an orderly presentation of CADET tools. It provides a standardized method for logging on the CADET system, and makes possible the accessing of tools and utilities through an on-screen menu, even though all the tools are not hosted on the same computer.

Both the novice and experienced CADET users needs were accommodated when the UIM was developed. After logging on the CADET system the user has two options to access the tools and utilities. The novice can select the menu approach, and is prompted through a series of menu-based screens starting with the main menu and ending at the desired tool or utility. The experienced user can select the direct approach and immediately access any desired tool, function or utility.

Reach Analysis Tool. The reach analysis tool used is the Crewstation Assessment of Reach, Version IV (CAR-IV) model (Ref. 3). CAR-IV is a design evaluation tool which aids designers in determining the percentage of the population accommodated by a designated crewstation geometry. It is an interactive tool hosted on a VAX 11/780. All data input and function selections are menu-driven, prompting the user to specify desired options and requested parameters.

CAR-IV consists of three program elements: the operator sample, the crewstation geometry, and the accommodation analysis. It utilizes two data bases, one for anthropometry data and the other for control/display locations. The user specifies the operator sample data (anthropometry) by either inputting actual body measurements or by generating a sample population using a Monte Carlo process. The crewstation geometry requires the user to specify the physical geometry (x,y,z coordinates and angle values) of the seat, crewstation controls, and the pilots position in the crew station. The accommodation analysis is performed using the data from the operator sample and the crewstation geometry. It positions each sample pilot in the crewstation and generates data on the following: ability to reach the controls, head clearance, and vision accommodation.

Workload Analysis Tool. The Human Operator Simulator (HOS) and the Human Operator Data Analyzer and Collator (HODAC) in combination provide a workload analysis tool (Ref. 4). Together they can produce a timeline analysis of operator tasks, a control/display usage analysis, and an operator loading analysis. The data bases used are control/display locations and aircrew tasks data. The HOS/HODAC models run on the Cyber 750 in the batch processing model. They are menu-driven and prompt the user to input data interactively. The HOS user must specify operator procedures and functions, the hardware procedures and functions, and the control/display layout in the crewstation. HOS simulates an operator functioning in the user specified environment. HOS simulates the perceptual, cognitive, and motor functions of the operator and the operating characteristics of the hardware. HODAC converts data generated by HOS into a suitable form useable by human factors analysts, and produces additional output reports and performs further analysis.

System Analysis Tool. Unlike the other CADET tools the system analysis tool, the Systems Analysis of Integrated Network of Tasks (SAINT)

software (Ref. 5), is not a prewritten model. SAINT is a network modeling and simulation technique which aides the designer in analysis and design of complex man machine systems. SAINT operates on the VAX 11/780 in the batch processing mode. Input data is entered interactively through the CADET workstation.

The designer creates a system in network terms using a symbol set. The symbol set describes tasks performed, precedence relationships among tasks, information flow through the system, and environmental stress effects on task performance. An analysis of the model developed is automatically performed by SAINT which provides summary information output concerning task performance, resource utilization, state variable status, and many other system performance measures.

Display Format Design Tool. The Display Format Design Tool allows the display designer to conceptualize symbol placement, information content, and color coding alternatives prior to formal display evaluations (Ref. 6). This tool utilizes the graphic mode of the Envision 230 terminal, the digitizing tablet with twelve button mouse, and the Envision 420 color printer, of the CADET workstation, as input and output devices. For permanent picture storage, the terminal is connected with the VAX 11/780 host computer and contains hard-disk storage.

The Envision firmware contains many useful two-dimensional graphic primitives including: lines, boxes, circles, polygons, and fill patterns. To simplify the communication task between the Envision terminal and the VAX computer several options have been developed, including: Draw, Restore, Save, Polygon, Grid, and Envision Utility functions. Graphics data can be entered using command sequences and by using the mouse and digitizing tablet to trace an object. The data base created is display formats.

The Data Base Manager. The Data Base Manager, as previously mentioned, is currently under development. The Data Base Manager will provide a way for the CADET system of tools to store, format, and facilitate the transfer of data from file to tool, tool to tool, and from file to file. This would eliminate the need to input duplicate data bases common to more then one CADET tool. Therefore, the control/display locations data base used by both CAR-IV and HOS would only have to be entered once as would the aircrew tasks data base used by HOS and SAINT. The Data Base Manager, when completed, will make the CADET system easier and more efficient to use.

VAX Utilities. The VAX Utilities provide an easy way for CADET users to create, modify, and maintain data files needed to run CADET programs. The VAX Utilities are accessible to the user through the UIM and provide file information, otherwise available only through the VAX DCL. Currently there are nine utilities: delete, directory, edit, print, purge, type, copy, diskspace, and rename.

FUTURE DIRECTIONS

One of CADET's key features is its modular structure which allows the system to grow and change with the user's needs. Several enhancements to CADET will make use of this feature. Some improvements on the drawing board are the addition of a Decision Aided Design System (DADS), tool enhancement, and the acquisition of more tools to broaden CADET's scope.

DADS will assist the designer by providing him with a human factors knowledge base. This knowledge base will make available to the user expert knowledge, which will help prevent him from violating human factor design principals. The tool enhancements include improvements to the reach analysis and system analysis tools.

The reach analysis tool will be improved by expanding the accommodation analysis output to include graphic representation of reach envelope overlaid on cockpit panels. Also, the range of seat back angle this tool permits will be expanded to allow for high back angle seat concepts.

The system analysis tool will be improved by the addition of a user-prompted method of data entry. This will allow the user to more quickly create a computer model of the crew system by which to evaluate design alternatives.

The workload analysis tool will be upgraded to include the mental aspects of crew procedures in the prediction of workload. Also, the tool will be structured such that empirical data from manned simulations may also be analyzed by this tool.

In order to increase the capabilities of the CADET system, other programs involved with computer aided design are monitored. Tools involving function allocation, mission analysis, and cost benefit analysis if acquired would enhance the capability of the CADET system.

SUMMARY

In order to remain a user-friendly system that is responsive to the designers' needs, a CADET system cannot be developed and left without continuing support efforts to maintain an effective, efficient system. For this reason, our CADET system, with its extensive user-friendly properties and diverse design and evaluation capabilities, can never be considered fully developed.

REFERENCES

1. Butterbaugh, L. and McBride, J. Lt.; A Computer-Based System for Workstation Design and Analysis; Conference Proceedings of Annual Meeting of The Human Factors Society; San Antonio, TX; October 1984.

2. Griffith, P.; Vax User Interface Module for the CADET System; BDM Corporation Report D-85-0225-TR; Dayton, OH; February 1985.

3. Harris, R. and Iavecchia, H.; Crewstation Assessment of Reach Revision IV (CAR IV) User's Guide; Analytics, Inc. Report 1800.10A; Willow Grove, PA; August 1984.

4. Glenn, F. and Iavecchia, H.; Human Operator Simulator User's Guide; Analytics, Inc. Report 1400.22C; Willow Grove, PA; March 1982.

5. Wortman, D. et. al.; Simulation Using SAINT: A User-Oriented Instruction Manual; AFAMRL-TR-77-61; Wright-Patterson AFB, OH; July 1978.

6. Griffith, P.; Display Format Design for the CADET System; BDM Corportion Report D-85-0221-TR; Dayton, OH; February 1985.

A Human Factors Approach to Display Specification

Gilbert G. Kuperman
and
Denise L. Wilson
Harry G. Armstrong Aerospace Medical Research Laboratory

Abstract

The determination of crewstation visual display parameters must be based on the task to be performed, the environment in which it is to be performed, and on the capabilities and limitations of the operator who is to perform it. The spatial, temporal, and luminance characteristics of the display may be specified on these bases. Only then may a meaningful trade-off analysis be performed against cost, size, weight, power, schedule and other, hardware-specific, display selection criteria.

Introduction

The selection of visual displays for integration into the work station is a frequently encountered problem for the human factors engineer. As in many such cases, a systems view of the problem can usually be expected to result in a more robust solution than can a point design approach. The operator will be invoked throughout this approach to serve as a check on the validity of each step. The method will be described in the context of a target acquisition problem, a task of recurring interest in military systems. Cathode ray tube technology will be assumed as the display medium.

Methodology

Information Requirements

The definition and quantification of information requirements is a two step process. First, the component parts of the target acquisition task must be unambiguously defined. Next the amount of information to be obtained about the target must be specified for each level of task performance.

In two current studies of crew workload associated with air-to-ground target acquisition, the authors have employed the following decomposition of the target acquisition process and have employed the accompanying definitions:

Detect: To determine the presence (absence) of an object of military significance
Classify: To determine the generic group to which an object

belongs (e. g., wheeled versus tracked vehicle)

Recognize: To specify the object's class (e. g., tank, truck)

(These definitions are similar to those invoked by other investigators as in the Target Acquisition Working Group [TAWG] Working Paper, 1972.) Ratches et al., 1975, employ a form of the Johnson Criteria (1958) in specifying the probability of correct detection and recognition for military vehicles as a function of the number of cycles across the target (where one cycle represents a pair of scan lines). An extract of these data follows:

Probability	Detection Criteria	Recognition Criteria
1.00	3.00	9.00
0.90	1.80	5.50
0.80	1.50	4.50
0.50	1.00	3.00

These criteria are based on targets encountered under conditions of low to medium clutter, where clutter is defined as objects of "target-like" appearance. Further, these data were derived under conditions of at least moderate target/background contrast and sensor signal/noise.

Sensor Parameters

While not necessarily of direct application to the display specification problem, several sensor parameters should also be considered by the human factors practitioner. These are:

Field of View (FOV): Navigation-related tasks generally require greater FOVs (and fields of regard) than do target acquisition tasks. Kuperman (1982) suggests that fields of view of about 20 to 30 degrees are typical of navigation sensors while much smaller angles apply to targeting sensors. In general, navigation information, especially if presented on a head-up or helmet-mounted, combiner-type display, should be at unity magnification with respect to the outside world while target information may be presented at greater magnification.

Active Scan Lines: The number of active scan lines is smaller than the total number of scan lines since a portion of the total scanning time is actually reserved for the vertical retrace interval at the display. Thus, a sensor specified to operate at 525 line rate will actually provide only about 485 active scan lines (Carlson, 1975). The angle subtended by an individual scan line may be found by dividing the vertical dimension of the sensor FOV by the number of active scan lines. This number must be further reduced by the Kell factor (typically 0.7) which accounts for the random orientation of the target pattern. Thus, the sensor can be treated as if it had only 70% of the number of

active lines or the number of lines on target can be increased by about 1.5 (Erickson, 1978).

Signal/Noise Ratio (SNR): Numerous studies have investigated the effects of SNR on operator performance. Gobel, *et al.*, 1980, review this literature and find mixed results. In general, performance improves with increasing SNR and SNR may be traded off against sensor/display resolution.

Operator/Display Parameters

Two correlated spatial parameters, display size and target subtense, two luminance-related parameters, gray shades and dynamic range, and one temporal parameter, critical flicker fusion, must be considered in specifying the display.

Angular Subtense: The angular resolution of the eye must be accounted for. A value of two minutes of arc per raster line is typical for non-laboratory conditions. If a target must be covered by 10 sensor lines for recognition, then it must subtend 20 arc minutes on the display. (Steedman and Baker, 1960.)

Display Size: Display size is a function of line rate, viewing distance, and the required number of lines on target. First, the target size on the display (d) is found by:

$$d = 2 \times \text{Viewing Distance} \times \text{Tan}(\text{Subtended Angle}/2)$$

For our example and a viewing distance of 30 inches, d = 0.175 inches. The vertical dimension of the display can then be approximated by multiplying the value by the ratio of the total number of active lines to the number of lines on the target (485/10). In our example, the display height would be about 8.5 inches. (If we assume a 4:3 aspect ratio, then the display width would be about 11.3 inches.)

Gray Shades: A gray shade is defined to be a square-root-of-two increase in display luminance. Two adjacent areas on the display, separated in luminance by a single gray shade, would be clearly distinguishable by the observer. The luminance of the lowest gray shade is measured with no signal applied to the display; it represents the reflected ambient illumination.

Dynamic Range: Dynamic range is defined as the square-root-of-two raised to the number of gray shades. Thus, a display exhibiting 6 gray shades would have a dynamic range of $(1.41)^6$, or 8:1.

Critical Flicker Fusion: The Ferry-Porter Law can be used to predict the point at which display flicker will no longer be discernable (to most observers):

12.5(assuming photopic adaptation) log(mean luminance) +37 Hz.

Summary Measures

Numerous figures of merit have been attempted in measuring display image quality. The modulation transfer function (MTF) forms the basis for a number of these expressions. All are based on the rationale that the observer must be able to distinguish target-internal detail. There is a minimum usable target modulation required by the observer. Modulation, M, is:

$$\frac{L_{max} - L_{min}}{L_{max} + L_{min}}$$

Task (1979) suggests that this value lies between two and five percent. Verona et al. (1979) assume a Gaussian-weighted writing spot on the CRT and compute the MTF of the display as:

$$MTF(f) = e^{-2(\pi\ \sigma\ f)^2}, \text{ where}$$

f = spatial frequency (cycles per display width)
MTF(f) = the display modulation at f
sigma = standard deviation of the Gaussian function

Take f = 485 (active lines) and limiting MTF(485) = 0.03. Then sigma = 0.000521 (in the same units as f). Using this value of sigma, the analyst can compute the remaining points in the MTF.

Task (1979) posits two cases of interest in testing the match between the display and and the operator: display-limited and operator-limited. He begins by noting that the required angular subtense of the target on the display must equal the product of the number of lines on target required for recognition and the required (by the eye) angular subtense of the individual display raster line. If this condition holds in actuality, then optimum use is made of the display. In the display-limited case, however, the raster lines actually subtend an angle much greater than that required. The result is that the target will subtend a larger than necessary angle on the display (since the number of lines required remains constant). In the observer-limited case, the individual raster lines are smaller than required by the eye and, hence, more lines will be required to achieve the required target angular subtense on the display.

Conclusions

A systems approach to display specification allows the human factors analyst to assure that the visual interface will be compatable with both the task to be performed and with the requirements/limitations inherent in the observer who is to perform them. An acceptable visual interface can be assured only

by treating the observer as an explicit part of the sensor/display chain.

References

Carlson, A. B., 1975, Communication Systems: An Introduction to Signals and Noise in Electrical Communication, 2nd. Ed., McGraw-Hill: New York.

Erickson, R. A., October 1978, "Line Criteria in Target Acquisition with Television," Human Factors, 20(5), 573-588.

Goble, L. G., Williams, L. G., Pratt, P. D., Wald, J., Rubin, G., and Hanson, T., March 1980, Human Operator Interface with FLIR Displays, AMRL-TR-79-114, Air Force Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Johnson, J., October 6-7, 1958, "Analysis of Image Forming Systems," Proceedings of the Image Intensifier Symposium, U. S. Army Engineer Research and Development Laboratories, Ft. Belvoir, Virginia. (AD220160)

Kuperman, G. G., July 1982, Systems Performance and Survivability Considerations for tactical Target Recognition, AFAMRL-TR-81-60, Air Force Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Ratches, J. A., Lawson, W. R., Obert, L. P., Bergemann, R. J., Cassidy, T. W., and Swenson, J. M., April 1975, Night Vision Laboratory Static Performance Model for Thermal Viewing Systems, Night Vision Laboratory, Ft. Belvoir, Virginia.

Steedman, W. C., and Baker, C. A., February 1960, Target Size and Visual Recognition, WADD-TR-60-93, Wright Air Development Division, Wright-Patterson Air Force Base, Ohio.

Target Acquisition Working Group, 10 August 1972, The Form of Visual Detection Data, Working Paper of the Joint Technical Coordinating Group for Munitions Effectiveness, Royal Aircraft Establishment and Naval Weapons Center.

Task, H. L., January 1979, An Evaluation and Comparison of Several Measures of Image Quality for Television Displays, AMRL-TR-79-7, Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Verona, R. W., Task, H. L., Arnold, V. C., and Brindle, J. H., A Direct Measure of CRT Image Quality, USAARL Report No. 79-14, U. S. Army Aeromedical Research Laboratory, Ft. Rucker, Alabama.

Users' Requirements for the Battlefield Management System

Carl Lickteig, Ph.D.
U.S. Army Research Institute, Fort Knox Field Unit

Abstract

This research investigated users' requirements for operating the Battlefield Management System (BMS) and the system's potential for downloading the small unit leader's workload. BMS, the primary product improvement planned for the M1A1 Block II modification, is an integrated complex of battlefield information acquisition, processing and communication technologies. While BMS is expected to significantly enhance command and control it may require a substantial change in the users' standard operating procedures, such as a shift from vocal (FM radio) to non-vocal communication modalities and from paper-and-pencil maps to electronic map displays. To ensure that users' requirements were included in the design of this innovative system an interface prototype was developed on an IRIS 1400 computer system. A cross-sample of small unit leaders, working from a digitized terrain data base (Fulda Gap) and menu-structured report functions, was required to construct and edit the map displays and tactical reports needed for their respective duty positions. Users' requirements for display features and control functions were obtained, as well as their recommendations for overall configuration, size and operating characteristics of the BMS interface.

Introduction

As the proponent of BMS, Fort Knox's Directorate of Combat Developments (DCD) has conducted a series of evaluations on BMS prototype systems to ensure that users' requirements will be included in the design and development of the M1A1 BMS, the primary product improvement planned for the Army's main battle tank. DCD's earlier efforts have determined the informational requirements with respect to mission accomplishment for small unit command and control. The current research identified users' BMS interface requirements, and future research is expected to focus on the BMS interactive requirements for intervehicular communication.

The BMS base display under evaluation was generated on the IRIS's 23 inch, high resolution monitor. Figure 1 depicts the overall configuration of this base display and the relative layout of the sub-displays. Users' interactions with all display features and functions were input by means of a tethered trackball and mouse device.

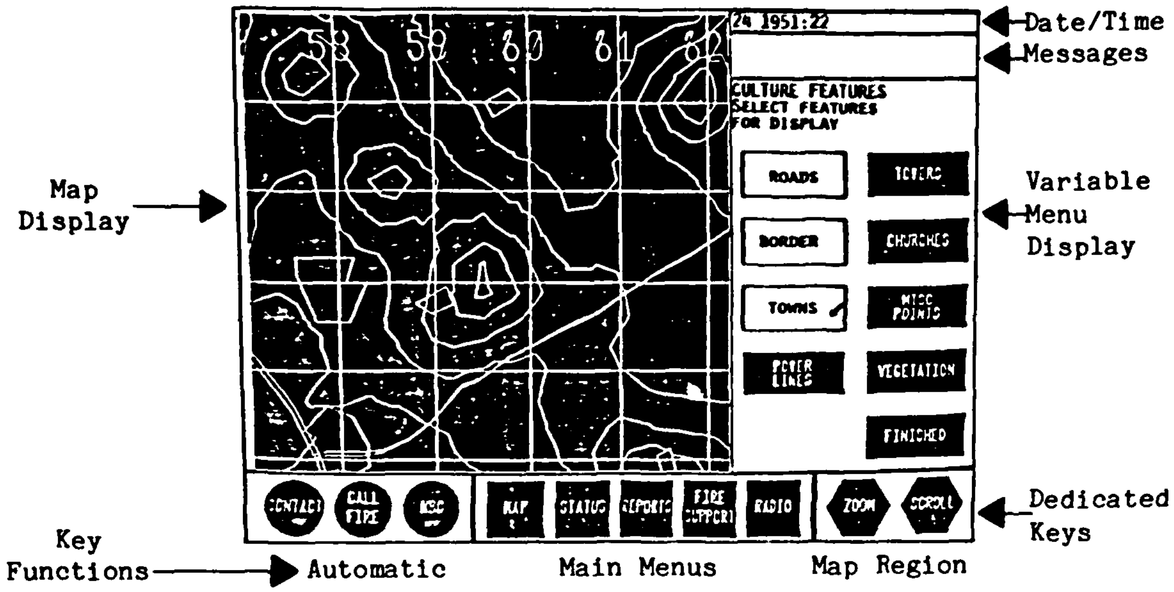


Figure 1. BMS prototype display depicting 5+ km area in the Fulda Gap.

The terrain display depicted in the BMS map window was generated from Defense Mapping Agency (DMA) Level I digital terrain of the Fulda Gap. Terrain and map features such as roads, towns, vegetation, grid and contour lines could be individually added or deleted to this map by way of menu selections. Figure 1 depicts the page within the BMS map menu's from which users were asked to select various man-made and natural terrain features (e.g., roads, border and towns). In addition a limited set of graphic control measures such as check points and tanks could be annotated onto the map area by either menu selections or a free draw function. Users were able to zoom in or focus on smaller topographical areas within the overall map region by menu selections that activated zoom, pan and centering functions. The variable menu area, provided users a display window for composing and editing a variety of reports. The actual "composition" was accomplished by users selecting the appropriate informational elements from a series of menu lists or pages.

The recent literature has provided numerous of guidelines for the design and development of user-friendly interactive computer systems. Working within these guidelines the BMS prototype was designed to test a variety of fundamental interface issues such as optimal display size and layout, achromatic vs. chromatic displays, free draw versus select, menu structure and organization.

Method

Subjects

Twenty-nine subjects were selected to participate in this evaluation. They came from three primary student groups at Fort Knox, Kentucky: 12 from the Armor Officer Advanced Course for company commanders; 11 from Advanced Noncommissioned Officer Course, platoon and first sergeants, and 6 from Basic

Noncommissioned Officer Course for tank commanders. These officers and enlisted soldiers were selected as a representative cross-sample of small unit leaders and as a basis for making comparisons about differences in users' requirements as a function of differences in training and operational background.

Procedure

During individual sessions (two hours) each respondent was placed in the role of "user" by providing him complete control over the BMS input devices, a mouse and trackball, and requiring him to construct his own map and report displays. Each soldier therefore began his session with a thorough familiarization and use of each of the map features and functions provided by the digitized terrain data base, and then proceeded to the report composition and editing tasks. As they viewed each of the prototype features and functions (cf. prototype description) users were directed to complete related items on the questionnaire.

Materials and Measures

The instruments used to obtain users' interface requirements for this evaluation were (1) a structured questionnaire and (2) a protocol record. The questionnaire was used to ensure that users' responses were collected for a uniform set of design issues, and the protocol record served as a backup measure for capturing any users' comments and recommendations not included in the formal questionnaire. The 65 items from the structured questionnaire addressed each of the BMS prototype features and functions operational at the time of the evaluation. A series of closed-ended items--rank order, dichotomous and 3 pt. Likert--developed for each cluster of BMS features and controls was followed by a related set of open-ended items.

Results and Discussion

This section discusses the findings for two of the most important clusters of BMS features and functions tested: the users' requirements to (1) call-up and manipulate a digitized terrain data base using the map features and functions and (2) compose and edit tactical reports in a modality independent of FM voice transmission.

Map Features and Functions

The transition from conventional paper maps to the electronic map formats marks a significant change in the standard operating procedures for all military personnel. A primary concern about the utility of the prototype's electronic map display was its comparability to the conventional medium of paper maps. To provide an index of comparability, users were requested to rank order ten (10) primary natural and man-made topographic features as portrayed by each medium with respect to their importance for armor operations. The five most important features selected by all users for both conventional and electronic maps were: contour lines, roads, vegetation, water and towns. While the rank orderings provide only a rough index of the comparability between conventional and electronic maps; they do suggest that the transition to electronic map displays may not significantly distort the users'

prioritization and utilization of map information. The protocol record also stresses the need to provide a BMS map format that is as comparable as possible to that of the conventional military map (i.e., feature portrayal and coloration should correspond to paper map formats).

Feature Selection/Deletion. One of the primary advantages provided by a digitized data based over conventional maps is that the former can be repeatedly "tailored" to the needs of different users as well as to the immediate task requirements that each user is currently executing. The user can selectively call-up any particular combination of map features and/or operational overlays relevant to the immediate course of action. Given the compressed map display area anticipated and tested for BMS (6 to 7 inches diagonally), users were unanimous in requesting that future BMS systems provide this selective call-up and delete capability.

Map Scale and Movement. The need for a more adaptable map format on BMS is further accented by the users' unanimous recommendations for each of the map control/movement functions provided by the prototype: pan, scroll and zoom. The protocol records, however, strongly suggest that the prototype's continuous zoom feature be coupled with a users' set of discrete zoom levels that automatically resolve to lower echelon areas of interest (i.e., 30 x 30 km, 6 x 6 km, and 3 x 3 km). Questionnaire items assessing the users' preferred map scales indicate that for the levels investigated (i.e., company commander, platoon leader, first sergeant, platoon sergeant and tank commander) a 1:50,000 is the most preferred, and a 1:25,000 is their second choice.

Pictures or Symbols. A central issue in this area of graphic representation is how best to generate a standardized set of tactical features for constructing operational overlays. Three out of four users requested that system-generated graphics use standard military symbols rather than pictorial images. Given the relatively minor system-memory requirements involved, however, it is recommended that all graphic features should be coded symbolically, pictorially and textually. By providing a textual label and definition to the user, especially the new user in cases of attrition, he could quickly review the information about the feature in question (e.g., the personnel and weapon system characteristics of a motorized rifle battalion).

Report Features and Functions

A reality of human communication is that speaking is a more fluent and informative modality than writing. BMS faces a serious challenge in its goal of bypassing the conventional voice mode of communication--the FM radio.

Menu Structure. Working within the confines of a dedicated report window rather than the entire BMS display area, the menu structures tested generally required users to (1) review and select from 7 to 10 "pages" of information to be included in their reports and (2) provide 14 to 20 key entries to complete a single report (e.g., SPOT, SITREP, NBC). Seventy-nine percent of all users described the menu structures designed and tested for this evaluation as "easy" or "very easy" to use and 89% reported that this extended set of report elements ensured that their final reports were complete and accurate messages.

Menu Modifications. Although the menu structure for report composition and editing appears to be a promising model for the BMS design; menu modifications for downloading the users' reporting workload might be to reduce either the number of reports included in the BMS software or to reduce the number of informational elements required to complete each report. Both alternatives were addressed by the questionnaire and a review of the complete results section provides BMS designers the users' preliminary guidelines concerning the frequency of occurrence for various reports and the relative importance of the informational elements doctrinally prescribed for selected reports. Protocol records also stress the need for additional BMS specifications related to the report function: system-generated wiring diagrams that clearly depict the structure of the hierarchical menus and a cursor-control dialog.

Integrated Map/Report Functions

One major conclusion, anticipated in the prototype's design, and confirmed by the protocol records, is that BMS must integrate the map and report functions. The real potential for BMS as a command and control system is to link spatial and verbal information, to synthesize map/graphic data with verbal/reports and orders. Nearly all military communications are based on the spatial geometry of the battlefield. This synthesis is not only critical to meet BMS expectations as a force multiplier, it is also essential to providing commanders a BMS that downloads the users' requirements. Two examples will be briefly discussed to illustrate this potential.

By integrating spatial and alphanumeric data at least some reports (e.g., SPOT and SITREP) could be completed on one "page" of the BMS display. "What" and "when" should be data automatically stamped to each BMS report when the report concerns a current activity. "Where" would be indicated by touching or positioning a cursor on the map display. "What" and "how many" could be entered from the same menu page if generic descriptors (e.g., heavy or light track, wheel, fixed or rotary air, and troops) were listed adjacent to a numeric keypad.

The final example, is the specification of automatic function keys such as Call For Fire (CFF), NBC, and Contact. To CFF, for example, the user might simply press the CFF key, touch the map area of enemy activity, and then press a "SEND" key. A request for suppression by indirect fire is then immediately sent to the artillery. Anyone familiar with the relatively complex and time consuming procedures required for target location (e.g., polar plots, shift from a known point) and the frequency of misdirected fires, can readily appreciate how important the integration of map and report functions is for downloading small unit leader job requirements.

In summary, this evaluation has provided a basis for the specification of a number of general BMS interface requirements and clear direction to numerous research and design issues that should be addressed more thoroughly. Finally, there is a need for objective measures of users' input patterns, speed and accurracy via BMS to compare against conventional or baseline workload requirements.

AN EMPIRICAL EVALUATION OF ANTHROPOMETRIC

ADJUSTMENTS FOR CREW SEATS

Russell Church and Sadie N. Decker Ph.D.

Martin Marietta Aerospace

Denver, Colorado

ABSTRACT

Little has been published about the comfort of multi-purpose crew seats. Our challenge was to write specifications for a seat that would provide the comfort of a casual lounge chair while providing support and safety in a tracked military vehicle. The operator of the vehicle will spend many hours on duty where workloads may range from sleeping and monitoring systems, to driving over rough terrain. This study employed an empirical approach for evaluating two adjustable seats, an Isringhausen and a Recaro. These two seats were tested against a typical non-adjustable vehicle seat which served as a control. Twenty-five Ai Force subjects served as evaluators. Analyses of the data indicated that in all anthropometric areas tested, the adjustable seats were significantly more comfortable than the control seat ($p < .01$). Additionally, the evaluation methodology is equally useful for testing a single seat.

INTRODUCTION

Most seats used in daily routines are compromised in design to accomodate the average sized user with a minimum of discomfort. However, the anticipated working environment of the crew cab being evaluated will require long hours of restricted movement where seating discomfort should not impinge on work performance.

The specialized requirements of a military vehicle frequently call for a seated crew member to perform a variety of tasks which may include driving over smooth or rough terrain, performing high-stress monitoring and communications tasks while parked, or spending long periods of inactivity, where the person is free to read, talk, relax or sleep. Typically, the operator of the vehicle will be confined to the seat except for brief and infrequent periods within the cab. Consequently, the seat must have a high degree of adjustability for operator size and be suitable to a wide range of activities.

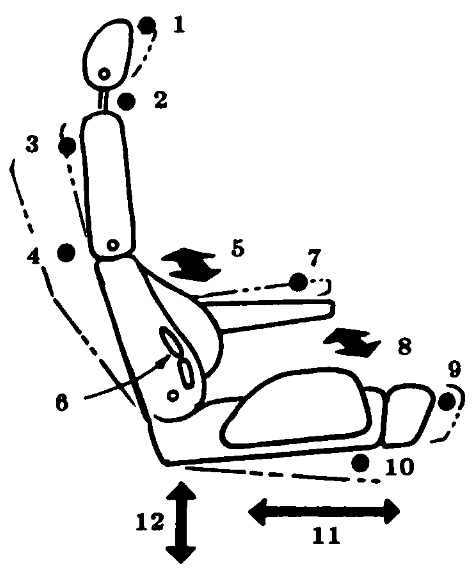
The intent of this research program was to perform a 3-phase test to evaluate the comfort and suitability of two adjustable test seats: 1) A 30-minute evaluation which focused on seat adjustments and comfort; 2) a longer evaluation which will involve dynamic human factors tests and simulated work conditions; and 3) a complete 24-hour simulated duty cycle. This report presents the results from the recently completed phase 1 study.

METHODS AND PROCEDURES

Experimental Design: Two adjustable seats, an Isringhausen and a Recaro, were compared to a third control seat which lacked any adjustment features. The three seats were compared for comfort, a sense of safety, and attractiveness.

Subjects: To obtain valid data which could be generalized to the population from which military vehicle operators might be selected, 25 evaluators (subjects) were referred to the study from Lowry Air Force Base, Colorado. Unit commanders were asked to provide both male and female subjects which included two females who approximated the 5th percentile in stature, and two 95th percentile males. As a result, the sample, which represents a cross-section of Air Force personnel, included 3 officers and 22 enlisted persons. Seventeen of the subjects were males and 8 were females. The height for the females ranged from 60 to 68 inches, and the males ranged in height from 68 to 75 inches.

Seat Adjustments: The first adjustable seat is manufactured by Isringhausen. It has several ergonomic accomodations and comfort adjustments. The second seat, made by Recaro, is an aftermarket seat for automotive use. It has more ergonomic accomodations and comfort adjustments (see Figure 1) than the Isringhausen. Exclusive with the Recaro seat is the adjustable thigh support (the length from the seat back to the back of the knee).



NO.	ADJUSTMENT	ISRINGHAUSEN	RECARO
1	Head Rest Fore & Aft	2 3/8"	2 3/4"
2	Head Rest Height	5"	2"
3	Upper Back Support	-	1"
4	Seat Back Recline	30° (partial)	85° (Full)
5	Lateral Back Support	-	3" (Across)
6	Lumbar Support	2 Air Cushions	3 Air Cushions
7	Arm Rest Angle	Infinite	*Infinite
8	Lateral Thigh Support	-	2 1/4" (Across)
9	Thigh Support	-	2 1/2"
10	Seat Pan Tilt	12°	-
11	Fore-aft	6"	8 1/4"
12	Height	** 1"	** 3"
13	Power Source	Pressurized Air	Elect. 12 VDC

* Added by the Space Operations Simulation Laboratory

** Not used in evaluation, substituted a footrest.

Figure 1. Adjustments for the Isringhausen and Recaro seats.

The control seat is a basic, inexpensive van seat. It has padded arm rests, a high seat back, and no adjustments. It follows the same principles of design (i.e., firmness, dimensions, and angles) that are usually employed in a non-adjustable van seat meant to accomodate all sizes of adults.

Testing Procedures: The two adjustable seats were located in a crew cab mockup which was fitted with a wall partition so that the subjects could see only the chair they were evaluating. This configuration was necessary to remove any visual bias that might influence the comfort ratings. To further ensure that attractiveness would not influence the comfort ratings, half of the subjects were assigned to the Isringhausen seat ($N = 12$) and half to the Recaro ($N = 13$) with one 5th percentile female and one 95th percentile male evaluating each chair. Also, the females were equally divided between the two chairs. Thus, two comparable, independent samples evaluated the adjustable seats. Each subject evaluated only one adjustable seat, however, all subjects participated in the control seat evaluation which was conducted in another separate crew cab ($N = 25$).

For the adjustable seats, a trained technician fitted the seat, footrest, and seatbelts to the subject's comfort. The subjects were told they would be seated for about 30 minutes (including adjustment time), and that they were not to stand up or re-adjust the seat. They were given a general interest magazine to read for the first 12 minutes. The magazine was removed and they remained seated for another 12 minutes. They were then escorted to a desk and asked to fill out a questionnaire which required them to evaluate the seat both objectively and subjectively. All subjects spent a similar amount of time in the control seat while performing a video game type task. The order of the seat evaluations was counterbalanced across subjects.

Rating Scales: The comfort rating scales were developed using principles derived from anthropometric studies of seat comfort (Drury and Coury, 1982). Pain/discomfort scales, as discussed by Shvartz et al., (1980) were not used because any well designed seat will not produce significant discomfort. Thus, the task was to develop a rating scale that would quantify subtle differences in body support for eight areas of the body which are known to be orthopedically related to job performance comfort (Diffrient, Tilley, & Bardagjy, 1983). A five-point rating scale was selected because it has the psychometric scaling advantage of increased sensitivity when compared to a two-point scale. The comfort rating scale was constructed as follows: 1 = poor, 2 = fair, 3 = satisfactory, 4 = good, and 5 = excellent.

RESULTS

Descriptive statistics were calculated for each seat using the comfort ratings for each of the eight anthropometric comfort zones. Figure 2 illustrates these mean ratings. When paired-sample T-tests were calculated comparing the questionnaire responses for adjustable seats to the control seat, the adjustable seats were rated as being significantly more comfortable ($p < .01$) in every category. With regard to comparisons between the Isringhausen and Recaro seats, the only significant difference found was in the area of lateral back support, which favored the Recaro, $F = 2.15$, $p < .05$. This had been expected because the lateral back support area is not adjustable on the Isringhausen seat. The Recaro seat scored significantly higher than the Isringhausen in overall attractiveness ($X^2 = 5.41$, $p < .05$). When asked to compare the comfort of the adjustable chair to a commercial airline seat, 69% responded that the Isringhausen seat was more comfortable, while 92% rated the Recaro as better.

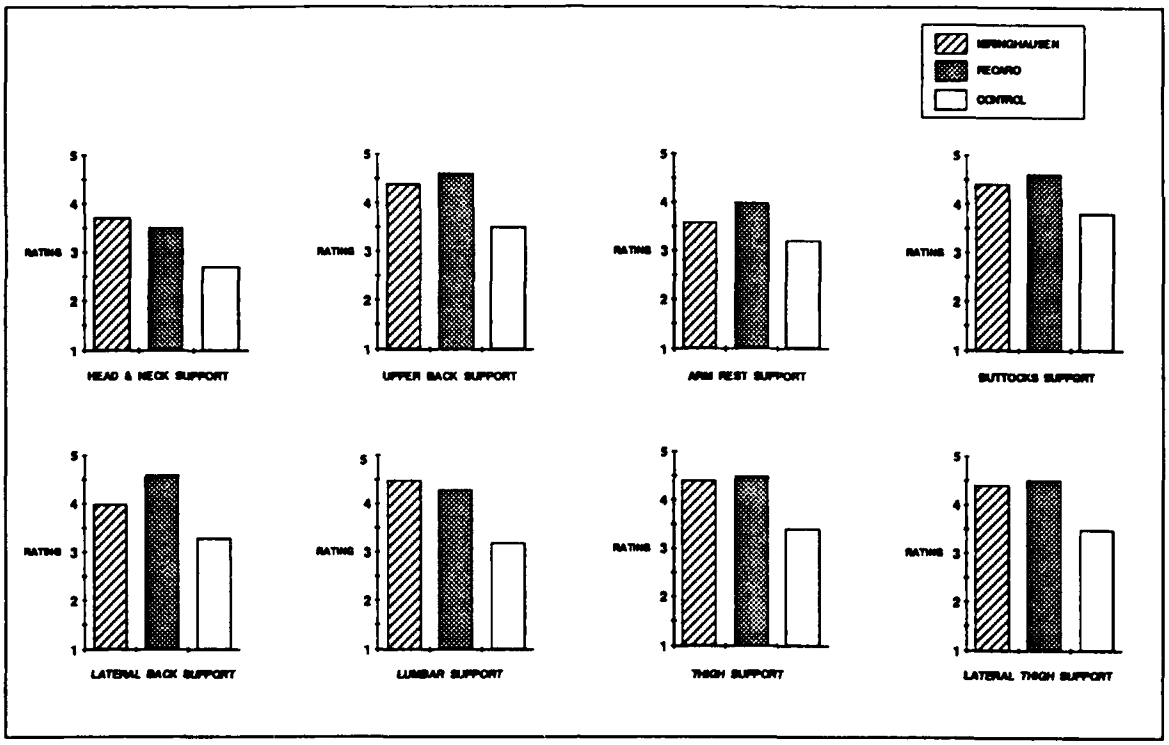


Figure 2. Mean comfort ratings for the Isringhausen, Recaro and Control Seats.

It is also important to note that the rating scale may be used for qualitative as well as for comparative evaluations. For example, the Recaro received a good to excellent rating for lateral back support while the Isringhausen was rated satisfactory to good. These qualitative differences may be important when long periods of sustained use are necessary. Using these scales, it will also be possible to compare a single seat to the data provided by this study.

DISCUSSION

We concluded that a seat with a high degree of adjustability is more comfortable than one with few adjustments. By shaping a seat to conform to the user's own body dimensions, proper support is provided which is more or less synonymous with comfort.

An important dimension not mentioned in this study is the seat pan height. This dimension is important because incorrect adjustment will result in too much buttocks pressure if the knees are too high and the thighs are not sharing this load, or too much pressure and reduced blood circulation in the lower thighs if the feet are not supported. The second phase of our research will address this issue.

Although it is difficult to relate attractiveness to performance, it is recognized that a desirable work environment enhances performance (Kleeman, 1981). The Recaro seat was clearly more attractive to our test subjects although such a requirement probably will not be a part of the human factors recommendations.

Finally, a technique which was piloted and which will be part of the long-term test, was that of video taping the entire duration of a subject's time in the seat. Because moving about in

one's seat is a semi-conscious effort to get more comfortable, the frequency of movement can be an indicator of discomfort. The number and type of movements will be compared to the user's evaluation questionnaire and will be further evaluated against his personal behavior characteristics.

ACKNOWLEDGEMENTS

We wish to express our gratitude to C. Ann Chaffee, the human factors lead for this project, and to Chris Schanker for his assistance in the analysis of the data. We also want to thank Jim Hylton and Cos Fazio for their expert editorial assistance and Margaret Glover for preparing the manuscript.

REFERENCES

Diffrient, N., Tilley, A.R., and Bardagjy, J.C. (1983). *Humanscale 1/2/3.* Cambridge: MIT Press.

Drury, C.G., Coury, B.G. (1982). A methodology for chair evaluation. *Applied Ergonomics,* 13.3, 195-202.

Kleeman, W.B. (1981). *The Challenge of Interior Design.* New York: Van Nostrand Reinhold.

Shvartz, E., Guame, J.G., White, R.T., Riebold, R.C. and Glassford, E.J. (1980). Effect of the circutone seat on hemodynamic, subjective and thermal responses during prolonged sitting. *Proceedings of the Human Factors Society* 24th Annual Meeting, 639-642.

Aiming Steadiness: Effects of Sex, Training, Weapon Weight and Aiming Position

Linda T. Fatkin and Gerald A. Hudgens
U.S. Army Human Engineering Laboratory, Aberdeen Proving Ground, MD

Abstract

Two experiments were conducted to investigate aiming steadiness in young men and women over repeated sessions. In the first experiment, subjects were asked to hold a pencil-like stylus in holes ranging in size from 3 to 13mm. Subjects were tested two days a week for two weeks. Steadiness errors were recorded automatically. The women were significantly steadier than the men, a difference which persisted over repeated test sessions. The women demonstrated improved performance at the most difficult levels, whereas the men's performance worsened.

In the second experiment, a modified pistol was used, with the stylus extending over the end of the barrel. Four weight adjustments were used and steadiness errors were recorded using two aiming positions. The women maintained their steadiness or aiming advantage over the men during the initial test sessions only. The men improved so that by the last session there were no signficant differences in performance. Applications to military situations, such as the aiming of hand-held weapons and training implications, are discussed.

Introduction

As we approach the 21st century, progressive changes in the make-up and requirements of the Armed Forces are reflected in personnel characteristics and in systems research and development techniques. Force readiness requirements, changing social attitudes, and new occupational patterns may dictate wider participation by women in more military skills and leadership positions than today. It is projected that all members of the Armed Forces will be potential combatants.

The implication of greater utilization of female personnel in the Armed Forces includes anthropometric considerations in equipment and weapons design and a renewed interest in the relative capabilities of male and female personnel on many performance variables. There will also be more reliance on laboratory studies and field experiments to measure the performance of military personnel within the weapons and training systems.

One area of concern which is appropriate for laboratory testing is the measurement of marksmanship skills. Early research studies have evaluated the basic skills used in successful weapons firing (i.e., manual dexterity and hand-eye coordination) through tests of hand steadiness. Attempts were made by Spaeth and Dunham (1921) and Humphreys, Buxton, and Taylor (1936) to evaluate the relationship between hand-steadiness performance and marksmanship. They found rank-order correlation coefficients of $r=+0.61$ and $r=+0.57$, respectively, between stationary steadiness and shooting ability. They concluded that, at least for men of known marksmanship

ability, involuntary movements consisitute an important and easily measured characteristic of successful marksmanship.

Although the early studies included only males who were experienced in rifle shooting, the hand-steadiness apparatus can be a valuable instrument for looking at other variables that may be relevant to the aiming steadiness of entry-level military personnel. Critical evidence concerning the effects of practice upon steadiness and the relation of steadiness to marksmanship would be afforded by testing both males and females who have had little or no shooting experience.

The first experiment described here examines gender differences in hand-steadiness performance over several training sessions. The second experiment examines the effect of additional variables, such as weapon weight and aiming position, on the steadiness of inexperienced males and females.

Experiment I

Method

Nine male and nine female military subjects were used in this experiment. The mean age for males was 23.3 years, and 24.1 years for females. Subjects were tested with their "preferred hand" which was described as the hand they usually used for writing.

Visual performance profiles were obtained using standard eye charts and an Ortho-Rater. No vision problems were identified.

The Standard Stoelting Co. Hand-Steadiness Tester used consisted of holes #1 through #9, ranging in size from 3 to 13mm. The position of the apparatus was adjusted according to each subject's arm length, so that a comfortable and consistent reach could be made. A shoulder support was adjusted vertically, so that the top of the hole apparatus would be in line with the top of the subject's shoulder, and horizontally (towards the subject's upper chest), to ensure that each subject would maintain an upright sitting posture throughout the testing. This portion of the apparatus rested lightly on the shoulders and did not provide any support to the arm during the testing.

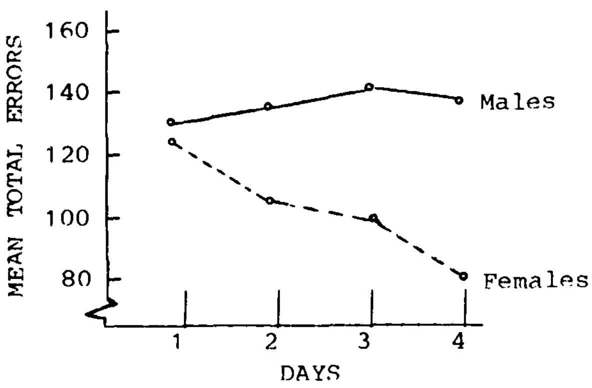
Using standard instructions, the subjects were asked to hold a pencil-like stylus in holes #4 (largest) through #9 (smallest) in that sequence. The number of times the side of the hole was touched in a 20-sec trial was recorded automatically. All subjects participated in two test sessions per day, with a break of at least 30 minutes between sessions. There were no reports of fatigue at any time. Each subject was tested two days a week for two weeks. Half were always tested in the mornings and half in the afternoons. Each subject was tested by a female or male experimenter on alternating test days to control for any possible sex-of-experimenter effect.

Results

A repeated measures ANOVA was performed with sex, test sessions, and hole size as the independent variables, and steadiness errors as the dependent variable. There were significant differences in the steadiness of men and women over the four test days, particularly in holes 7 through 9, $\underline{F}(15, 240) = 2.234$, $\underline{p}<.01$. As demonstrated in Figure 1, the women's performance improved over the four test days (fewer errors), while the men's errors increased.

Figure 1

HAND STEADINESS
EXPERIMENT I (N = 9/GROUP)



Experiment II

Method

Twelve male and twelve female college students, all inexperienced in weapon firing, participated in this study. The mean age for males was 18.4 years, and 23.1 years for females. Subjects were tested with their preferred hand on the hand-steadiness apparatus described in Experiment I, and were also tested for visual acuity.

Data were analyzed for errors recorded for 10-sec trials in holes 1, 3, 5, and 7. In this study, the stylus extended from the barrel of a hand-held weapon (the modified frame of a .45 pistol). Four weapon weights were used, ranging from the lightest (#1) which was comparable to the M15 .38 Revolver with one round, to the heaviest (#4) which compared with a fully-loaded M1911A1 .45 Pistol.

Two standard aiming positions, supported and unsupported, were used. In the "supported" position, subjects were instructed to hold the weapon with their preferred hand and maintain a steady-hold position by supporting it with their non-preferred hand. In the "unsupported" position, subjects held the weapon with their preferred hand only.

As in Experiment I, subjects were tested two days a week for two weeks, with no reports of fatigue. Half of the subjects were tested by a female experimenter and half by a male experimenter. The order in which the four weapon weights and two aiming positions were presented was counterbalanced across subjects.

Results

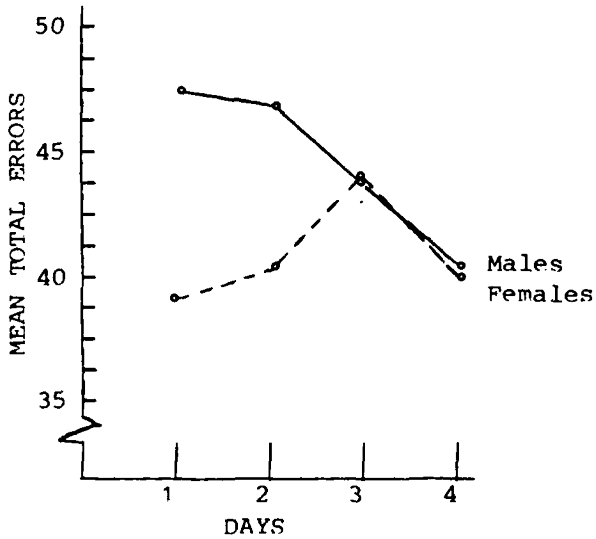
A repeated measures ANOVA was performed with sex, weapon weight, aiming position, test sessions and hole size as the independent variables, and steadiness errors as the dependent variable. Significant differences in aiming steadiness were found between weapon weights across all subjects, $\underline{F}(3, 66) = 26.99$, $\underline{p}<.01$. Subjects were steadiest with weapon weight #2 (which is approximately the weight of a Smith & Wesson 9mm Pistol with one round), and made the most errors with the heaviest weapon, #4.

Subjects also made twice as many errors using the unsupported aiming position than when using the supported aiming position, $\underline{F}(1, 22) = 219.94$, $\underline{p}<.001$.

Unlike Experiment I, the women maintained their aiming steadiness advantage over the men during the initial test sessions only, $F(3, 66) = 3.10$, $p < .05$. As illustrated in Figure 2, the men improved so that by the last session there were no significant sex differences in performance.

Figure 2

AIMING STEADINESS
EXPERIMENT II (N = 12/GROUP)



Discussion

In Experiment I, the gender and practice variables seem to have the greatest effect on steadiness performance, while in Experiment II, weapon weight and aiming position, in addition to gender and practice, are the major factors. Although the overall pattern of sex differences in aiming steadiness was different in Experiments I and II, the data from both experiments support the importance of including repeated measures in the experimental design. Caution should be exercised in drawing conclusions regarding male/female performance differences based solely on initial test trials, as a small amount of experience or training can have a dramatic effect on the relative performance of the groups.

The use of a laboratory approach for skill development in entry-level military personnel fits in well with the more recent revisions of military training programs where simplified fundamentals are being stressed before moving on to field firing exercises (RM, Trainer's Supplement #2, 1980).

This approach can also provide valuable data in the area of weapons systems design. For example, the results from Experiment II indicate that aiming steadiness was effected by weapon weight. A weapon that is too heavy may increase the chance of aiming errors, yet one that is too light may not sufficiently calm the natural hand tremor present in most individuals.

Future efforts will focus on more applied research efforts which would evaluate the relationship between aiming steadiness and marksmanship ability. Improving marksmanship by initially training for steadiness with a laboratory apparatus instead of actual field firing would support recent

efforts of military cost and training effectiveness analysis (CTEA) of several weapons systems (Smith, Osborne, Thompson & Morey, 1980). The intent is to find training procedures and devices that produce maximized personnel/system performance at an affordable cost. The ultimate expectation is to provide a natural transfer of skills acquired through laboratory tests and/or field experiments (e.g., better hand-eye coordination) to the more complex task of engaging targets at range.

References

Humphreys, L. G., Buxton, C. F., & Taylor, H. R. (1936). Steadiness and rifle marksmanship. Journal of Applied Psychology, 20, 680-688.

RM, Trainer's Supplement #2 (1980). Fundamentals of Rifle Marksmanship. Washington, DC: Headquarters, Department of the Army.

Smith, S. Osborne, A. D., Thompson, T. T., & Morey, J. C. (1980). Summary of the ARI-Benning Research Program on M16A1 Rifle Marksmanship (Research Report 1292). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Spaeth, R. A., & Dunham, G. C. (1921). The correlation between motor control and rifle shooting. American Journal of Psychology, 56, 249-256.

Sustained Human Performance in the Heat:
Summary of an On-Going Systematic Research Program

Bernard J. Fine and John L. Kobrick

U.S. Army Research Institute of Environmental Medicine, Natick, Massachusetts

Abstract

In a recent review of the effects of climate on performance, we found nearly 50% of the climate-performance research to be deficient when evaluated against a set of rigorous criteria. It was virtually impossible to make definitive statements about the effect of climate on performance. In this paper, we summarize results to date of our program of research on the effects of heat on sustained performance of military tasks, designed to satisfy the aforementioned criteria. A remarkable consistency of baseline data over five studies is shown. Within the confines of the performances we are studying, a data base is evolving which should enable prediction of performance at various temperatures of groups of soldiers clad in battle dress uniform or chemical protective clothing.

Introduction

In a recent review of the literature on the effects of climate on human performance (3), we found that nearly 50% of the research was deficient when evaluated against a set of criteria which included (a) number of test subjects (Ss), (b) adequacy of training on performance tasks, (c) appropriateness of research design, (d) appropriateness of statistical analyses, (e) clarity of presentation, and (f) equipment-related problems. We concluded that it was impossible to make definitive statements about the effect of climate on performance and noted three major shortcomings in the research: (1) reliance on unrealistic and irrelevant tasks for assessing performance, (2) inadequate training on the tasks and (3) insufficient length of exposure to the climatic stressor while performing the tasks.

In this paper, we present the results of our own program of research on the effects of heat on human performance. In addition to trying to satisfy the research criteria noted above, we have attempted to facilitate direct comparison between the studies in our program by using similar methodologies, identical performance tasks and identical baseline environmental conditions from study to study. There have been five studies to date, designated herein as Studies I, II, III, IIIA and IV.

Method

Each of the five studies had the same nucleus of "basic" tasks, the same study design and used different Ss, with the exception of Studies III and IIIA which shared the same Ss.

Tasks

The tasks, which included aspects of those performed by members of

artillery fire direction centers (FDC's), forward observers and Army communications personnel were:

1. Computation of "Site:" - "Site" is an aiming adjustment used by FDC's when firing. Tape recorded data were transmitted to the Ss over headsets in a format similar to that in an FDC. The men logged in the data, performed arithmetic calculations, entered data into and read answers from an FDC slide rule and wrote the answers. In this task only, each message was repeated once.

2. Receiving and decoding map grid coordinates: - Pre-recorded, coded coordinates were transmitted as radio messages via headsets. The men logged in the alpha-coded coordinates, chose the correct one of three decoding devices, translated the code into numeric format and recorded the answer.

3. Receiving and decoding messages: - Pre-recorded, coded messages of from 5-8 words were transmitted in military format over headsets. The men logged in each message, decoded it (codebook) and recorded the transcription on a form.

A fourth task, map plotting, was used in Studies II, III, IIIA and IV. Each S had a list of grid coordinates and a map with pre-plotted battery locations. S plotted targets (Studies II, III, IIIA, IV) and, also, found the ranges (Studies III, IIIA and IV) and deflections (Study IV) of the targets from designated batteries.

Ss did not know which of the three radio message tasks to perform until a message arrived. Messages were of military format and included a variety of voices and transient background noises. Several other tasks were performed during the four studies, but they will not be discussed here.

Design and Procedure

Ss arrived at three-week intervals (two-weeks for Study I) in six-man groups. Each group finished its assignment before the next arrived. The groups had two weeks of training (one week for Study I) followed by an "experimental" week to evaluate performance in the particular environment(s) being studied.

Ss trained in a classroom setting, six to seven hours a day. Training on the codewheel, codebook and Site tasks began with a simple written format and increased in complexity until Ss could handle the rapid, noisy radio messages. Ss practiced several hundred messages with feedback and discussion of errors. Accuracy initially was stressed over speed of performance. Map-plotting (Studies II, III, IIIA and IV) was practiced for hundreds of trials with immediate feedback and correction of errors. Again, emphasis at first was on accuracy rather than speed.

The design for the "experimental" week was similar for all studies. On Monday, two hours of practice re-established pre-weekend performance levels. Tuesdays and Thursdays, control days to establish baseline performance, always were at 70°F.,35% rh. Wednesdays and Fridays were "experimental" days and their environments varied. (Studies I and II each included an exposure to altitude and Study IV included exposure to a "cool" condition with Ss clad in chemical protective clothing. These experiments are not discussed here.) Exposure time always was seven hours/day. Clothing sometimes differed from

study to study. The purpose, control and experimental conditions and clothing worn in each study is shown in Table 1.

Table 1

Study Number	Purpose	Control Conditions	Experimental Conditions
I*	Effect of heat on performance	70°F; 35% rh fatigue uniform	heat, 91°F; 91% rh fatigue uniform
II	Effect of heat on performance	70°F; 35% rh fatigue uniform	heat, 97°F; 73% rh fatigue uniform
III	Effect of heat on performance	70°F; 35% rh fatigue uniform	heat, 105°F; 40% rh fatigue uniform
IIIA	Effect of heat and rapid translocation on performance	70°F; 35% rh fatigue uniform	heat, 105°F; 40% rh plus simulated translocation fatigue uniform
IV	Effect of heat and chemical protective clothing on performance	70°F; 35% rh battle dress uniform (BDU)	heat, 91°F; 61% rh **BDU + MOPP IV

* This study has been published (2).
** MOPPIV = Chemical Protective Clothing, complete ensemble.

The "radio" tasks were presented as one-hour blocks of messages, 25 per hour. Intervals between messages ranged from 30 seconds to over two minutes. The same messages were used in all studies. The messages were presented during hours 1,3,5,7 on the control and experimental days. Concurrent with working on messages, Ss worked on the map task (except in Study I where they simply waited for messages). Radio messages always had priority; map work stopped upon hearing a message, to be resumed only when message task was completed. During hours 2,4 and 6, Ss worked only with maps (Studies II,III,IIIA,IV), at their own pace for 35 minutes without interruption. They were tested on other tasks during the remaining time. (In Study I, in which there was no map task, the men monitored infrequent radio messages, three or four messages per hour, in order to test "alertness.") A ten-minute rest break was given each hour.

Results and Discussion

Results are presented only for Site calculation, the most difficult task. Error rates for this task were lower than for the others, presumably because its message was the only one presented twice. Figure 1 shows the control values (percent group error by elapsed hours of work) for each of the five heat-associated experiments. The higher error rates in Study I probably are due to the shorter training time. The lower error rates in Study IV may be due to the notably higher caliber of S found in that study and/or to the Ss being

especially motivated to perform well in a study of critical chemical protective gear. However, the important point to note is the very even performance of each group in each of the control conditions over a seven hour period. This is not typical for studies of this kind and attests to the value of long, intensive training periods.

For each of the five heat-related experiments, the raw error scores were transformed into percentages of the maximum possible error obtainable for the Site calculation task. Difference scores were then computed between the individual control error scores and the error scores obtained when the task was performed in the heat. The difference scores then were averaged across Ss for each study and are presented as percent mean group error in Figure 2. The curves in Figure 2 thus show the decrements in group performance over time which occured in the various heat stress conditions with each condition compared with its own control. In effect, Figure 2 is the beginning of a nomograph which may eventually allow us to predict group performance on various tasks for troops who will be exposed to selected environments for given periods of time. What differentiates this nomograph from others is that this one is not based on a collection of data from unrelated studies with different methodologies, invalid tasks, varying durations of exposure and, often, poorly trained subjects, but on information derived from a common data base, with relevant tasks performed by well-trained Ss for longer periods of time than heretofore has been the case.

The precision of the data in Figure 2 is worth mentioning. While we have noted previously the limitations of Effective Temperature as an index of stress (3), it, nevertheless,can be used for purposes of illustration. Note that the impairment of the group wearing MOPP IV at 91°F;61%rh (Study IV) falls between that of the groups performing at 95°F;91%rh (Study I) and 97° F; 73%rh (Study II), even though the heat stress in Study IV was less severe than in either Study I or Study II. The Effective Temperatures for the three studies are 92 (Study I), 89.5 (Study II) and 81 (Study IV). However, Ss in Study IV were wearing chemical protective clothing and a "rule of thumb" is that the protective clothing adds about ten degrees to the ambient temperature (1). Adding ten degrees to the effective temperature of Study IV raises it to 91, intermediate between the 92 and 89.5 values of Studies I and II and precisely where its performance decrement falls. This at once supports the "rule of thumb" and the validity of our "nomograph."

References

1. Breckenridge, J.R. U.S. Army Research Institute of Environmental Medicine, Natick, MA. Personal Communication.

2. Fine, B.J. & Kobrick, J. L. Effects of altitude and heat on complex cognitive tasks. Human Factors, 1978, 20, 115-122.

3. Kobrick, J.L. & Fine, B.J. Climate and Human Performance. In D.J. Oborne & M.M. Gruneberg (Eds.), Psychology and Productivity at Work: The Physical Environment. Wiley: London, 1983, pp. 69-107.

Control Values for Five Heat Studies

Site Calculation Task

Study I N=28	Study II N=22	Study III N=20	Study IIIA N=20	Study IV N=20

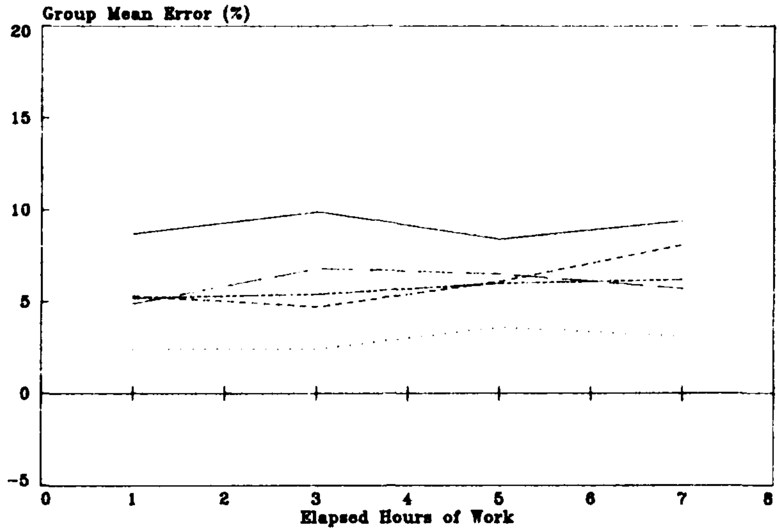


Figure 1

Effect of Heat on Calculation of Site

Heat Performance Scores - Control Scores

Study I 91°F;91%rh	Study II 97°F;73%rh	Study III 105°F;40%rh	Study IIIA 105°F;40%rh	Study IV 91°F;61%rh

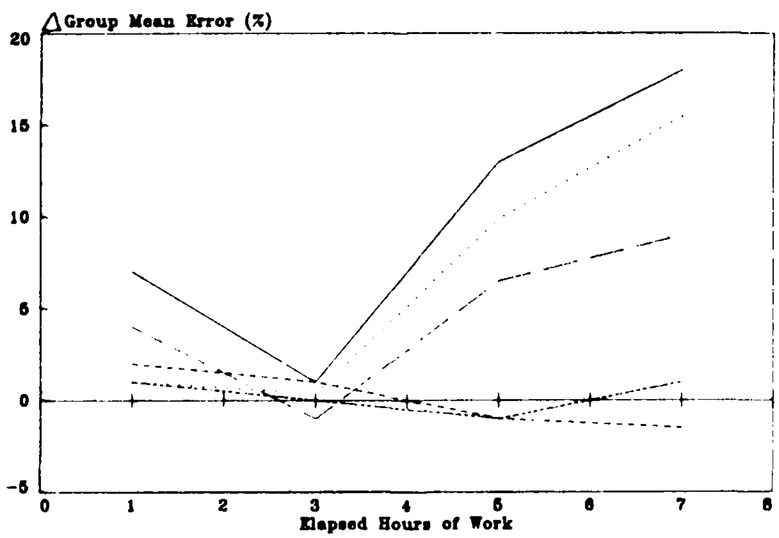


Figure 2

Estimating Operator Performance in NBC Conditions

David J. Travale
Valerie J. Gawron, Phd.
Arvin/Calspan Corporation, Buffalo, New York

Abstract

A human performance model was developed by Calspan as part of the U.S. Air Force CAT phase I program. The major objective of the model, called HUMAN, was to provide a tool which would be useful in simulating a wide variety of human activities. HUMAN can model these activities by breaking them down into specific tasks which are networked to detail an entire activity. A major advantage of such a model is that human performance under NBC conditions which is not easily measured in the real world can be simulated using HUMAN.

Introduction

Background

In today's world of ever increasing non-conventional warfare threats, there is a large problem in determining how personnel will perform under NBC conditions. For military commanders to maintain maximum effectiveness, they must know what tasks humans can perform after exposure to an NBC environment. This information is also critical to hardware and software engineers when designing equipment which is to be operated under NBC conditions.

When considering human performance in the NBC environment, there are three types of situations which must be dealt with. The obvious situation is when a harmful agent is introduced into the environment by aggressor forces. Under this condition death or serious physical impairment may occur to personnel causing mission degradation. The second and not so obvious condition is when personnel intentionally take chemicals as protection against the previously mentioned harmful agents. This situation, although not as critical as the first, may also degrade mission performance as a result of the particular drug taken. The third condition involves the wearing of chemical protective clothing. Operator performance while wearing protective gear may be degraded for a number of

reasons ranging from vision impairment to heat stress.

In order to have full control over combat situations, commanders need to know the limitations of their personnel to make sound decisions. To determine these limitations it is not always feasible to run large scale experiments, and in the case of harmful agents it is unethical to expose personnel to NBC conditions. This is where HUMAN would be used to get an estimate of the operator performance under the NBC conditions.

Discussion

HUMAN is an acronym which stands for Human Utilization Model and Analytic Network. The user can select the conditions to define the task to be performed by the human operator. Large tasks can be modeled by using the networking capability of HUMAN. This is achieved by breaking the overall task into subtasks corresponding to each node in the network.

The conditions of each node in the network are separated into three categories called vectors. Using these vectors, the user can input values to model what a human may encounter. The vector descriptions are as follows:

Vector 1 - 'Environmental Status'
This vector contains variables which describe the operator's surroundings. Examples of these are: temperature, humidity, and ensemble conditions.

Vector 2 - 'Operator Status'
This vector contains variables which describe the operator's physical capabilities and present condition. Examples of these are: time on task, drug taken, and drug dosage.

Vector 3 - 'Task Variables'
This vector contains variables which describe the type of task the operator is to perform. Examples of these are: number of displays to monitor and number of alternative responses.

HUMAN then requires the user to input the skill(s) the human will need to perform the desired task. A menu consisting of 10 skills will be given for the user to assign weights to. The total weight for all 10 skills must sum to 1.00. Therefore, should a task require detection and tracking skills with equal importance, both of their assigned weights should equal 0.50.

The output of the model is estimated: time to complete the task, probability of completing the task, and percentage of errors incurred while performing the task. Additional features of the model include sensitivity analysis across an independent variable as well a capability to plot the results.

A key element of HUMAN is a human performance data base containing records of empirical studies. The data gathered can include actual experiments in the case of determining performance while in chemical protective garb, data gathered as a result of accidental exposure to high radiation, and so on. Each of the records contain the human performance data from the studies along with the conditions under which the study was performed. For example, the conditions might be that the subject was performing a visual detection task, the response mode was tactile, and the subject had taken 30 milligrams of pyridostigmine. The performance data may be the time to perform the visual recognition task as a function of the time since the drug was last taken (Graham and Cook (1984)).

The data, including the conditions, is then coded into the corresponding subroutine of HUMAN. Thus, if a modeler using HUMAN selects conditions which match the conditions of the coded study, then and only then is the performance measure value for that study selected as output. If the conditions of several studies are met, the program will apply a weighting system to the performance measure value of each study. Using the weighting system, a single weighted average performance value will be generated by the subroutine. This weighting is determined by the number of independent variables that experimenter controlled. Thus the more controls placed on the study, the greater the influence that study had when performance values were averaged. In general, it would be expected that the more variables monitored during an experiment, the greater the accuracy of the results.

Example

Assuming we wanted determine if taking 30 milligrams of the drug pyridostigmine has any affect on an operator's ability to perform a detection task. Using HUMAN, the user would select to model a single task. Next the vector elements must modified, specifically the operator vector. The element, drug type, of the operator vector is set to pyridostigmine. The element, drug dosage, is set equal to 30 milligrams. Now the user must set the skill weighting for a detection task to 1.0. This would indicate that detection is the only skill we are interested in. Finally the user would select the dependent variable time as a performance measure and then run the model. For the above element values, HUMAN estimates the detection task to take

6.8 seconds.

To compare operator performance when not taking pyridostigmine, the user sets element 9 to placebo. All other element values and skill weights will remain the same. Under these conditions, HUMAN estimates the time to complete the detection task as 6.6 seconds. Since there is only 0.2 seconds difference between the two conditions, it is reasonable to say pyridostigmine has little effect on a person's ability to perform a detection task.

Concluding Statements

The human performance model allows a user to model large tasks and receive as output an estimated performance of those tasks. All performance measures output by HUMAN are the result of empirical studies which can be easily traced to the original hardcopy report for verification. The main advantage is that users unfamiliar with a subject area can easily test the effect of an independent variable on task performance.

Planned future model enhancements include application of artificial intelligence techniques to aid both the novice and experienced user.

References

Graham, C., and Cook, M.R. (1984, August). Effects of Pyridostigmine on Psychomotor and Visual Performance. (Technical Report AFAMRL-TR-84-052). Wright-Patterson Air Force Base, Ohio: USAF Aeromedical Research Laboratory.

A Rapid Entry/Egress System for Chem Defense Shelters

Valerie J. Gawron
Saverio M. P· ˙iese
Charles K. Akers
Calspan Corporation

Abstract

The protection of air and ground crew, soldiers, and medical personnel in a nuclear, biological, or chemical (NBS) contaminated environment is currently provided by protective garments and collective shelters. The current method of entering or exiting these shelters is through an airlock system. This process may take as long as 15 to 20 minutes to complete. To dramatically reduce the entry/exit times as well as to eliminate the need for consumables associated with airlock technology, Calspan Corporation has developed a unique inter-mechanism. This mechanism uses zipper-like fastener technology to provide a method for the rapid movement of personnel between protective garments and shelters. An operational description of this novel concept with specific applications is discussed.

Introduction

Current Problem

Current safeguarding of military personnel in a NBC contaminated environlent uses full protective garments and collective-protection structures. Indivudal protection is accomplished by encapsulating the personnel in a garment that is impermeable to toxic vapors and particulates. Respiration and ventilation are provided which remove the hazardous material by mechanical filtration and absorption. However, the personal-protection garment has few or no mechanisms to allow full intake, performance of toilet functions, in-situ emergency medical attention, or rapid transfer of the wearer from the protective garment to another mobile or stationary protective environment. Therefore, personnel can perform field activities for only a limited duration. Collective shelters, on the other hand, are rigid or semi-rigid structures that permit long-term personnel activities. Examples include stationary systems such as field hospitals, communication centers, aircrew-ready rooms, and mobile systems such as troop transports and armored vehicles. These systems can provide a safe environment without the encumbrance of the personal-protection garment.

The need exists to rapidly move personnel, litters, and materiel into and out of a collective shelter without exposing the interior environment of the shelters to the risk of chemical contamination. There are also requirements

for the personnel stationed within the shelter to exit and enter the shelter, it is required that all entry and exit be made through an airlock system. This system must be adequately ventilated and of sufficient size to accommodate decontamination, don/doffing, and storage of the protective garments.

A single entry or exit event may take from 15 to 20 minutes to complete. Moreover, this method imposes additional demands for consumables such as decontamination agents and for additional air handling and filtration capacities. The current system imposes a tremendous burden on the overall effectiveness of the personnel exiting or entering the shelter. A need exists to dramatically increase the flow of personnel and supplies into and out of the shelter while maintaining a safe, "shirtsleeve" working environment for the personnel and support staff within.

Potential Solution

Calspan Corporation has recently developed, under Internal Research and Development funding, an Interchange Mechanism that provides a methodology for the rapid transfer of personnel and materiel into and out of collective shelters and other clean environments. This mechanism, which uses conventional zipper-like fasteners, can greatly reduce the time and material requirements for entry and exit. Entry/exit times of less than two minutes have been demonstrated for the personnel/shelter application using prototype hardware. The device also has application to other NBC-environment requirements, such as transfer of flight personnel to and from aircraft, the handling of the soldier's body functions and taking of nourishment, and medical casualty handling.

Calspan Interchange Mechanism concept permits the coupling of zippers on a shelter and another protected enclosure, such as a garment, as shown in Figure 1. After the zippers are coupled, movement of the Interchange Mechanism results in the following actions within the body of mechanism to occur: (1) the zippers are opened, (2) the individual zipper tapes are rotated through 90 degrees, and (3) the zippers are reconnected so that a passageway is created between the interiors of the garment and the shelter. The interchange mechanism comprises four distinct components (see Figure 2). The upper section consists of two separate zipper sliders: one is affixed to the assembly and the other is free to operate in the conventional manner or may be locked to the assembly to operate as part of the unit. The central portion accepts the separated ends from each unfastened zipper, rotates them through 90 degrees, and provides the reoriented components for interconnection in the bottom section. The lower portion of the mechanism is a single unit which permits opening (or closing) of two parallel fasteners simultaneously; it consists of essentialy two conventional zipper slides. In operation when the Interchange Mechanism is moved up the zipper tracks, the upper slides separate the zipper tapes, the central portion rotates the tracks, and the lower slides interconnent the rotated components in a single motion. The procedure is, of course, reversible. In most applications of the mechanism, the smaller of the components would be attached to the protective garment to minimize weight or bulk inconvenience for the encapsulated person. The larger component would be located on the less mobile system, such as the collective shelter.

In a recent full-scale concept demonstration, the Interchange Mechanism was successfully used to transfer personnel from a chemical-protection garment to a simulated collective-protection shelter. This simulated entry was effected in less than two minutes:

- despite the prototype nature of the interchange and engagement mechanisms,
- without optimization of zipper tooth (scoop) geometry or deployment configuration, and
- without prior rehearsal.

Thus, the Calspan Interchange Mechanism clearly provides the technology for the rapid transfer of personnel and materiel between protective environments.

Specific Applications

To illustrate the universal nature of the Interchange Mechanism, a number of specific applications are shown in the following figures. These illustrations indicate how the Interchange Mechanism can support ground and air crew, the soldier, and the medical personnel for operational readiness in a contaminated environment.

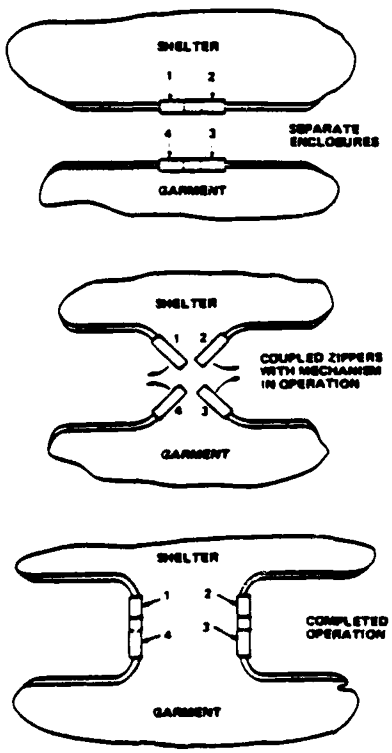


Figure 1- Schematic Diagrams Illustrating Functioning of Calspan Interchange Mechanism (Numbered Elements are One Zipper Tapes)

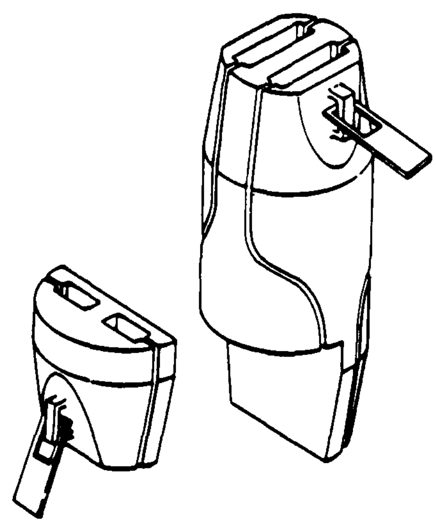


Figure 2- Assembled Interchange Mechanism

Figure 3- Nourishment for Encapsulated Troops Can Be Provided in Protective Pouches that are Zipped onto Outer Garment with the Interchange Mechanism, Similar Pouches Can Be Used For Urination.

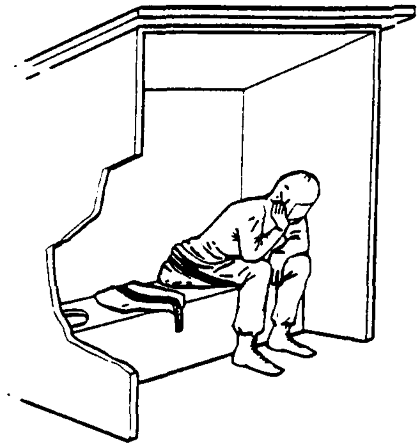


Figure 4- The Always Perplexing Problem of Providing for Toilet Functions of Protected Troops Can Be Solved Either with Disposable Zip-on Pouch or with a Suitably Designed Attachment to a Conventional Toilet.

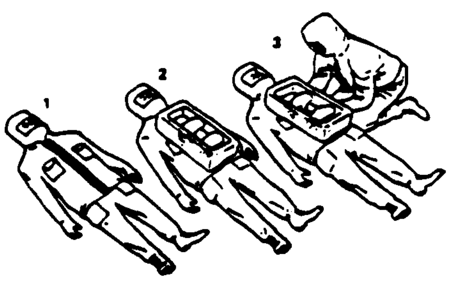


Figure 5- Medical Glove Box for In-the-Field First Aid. Current Model is Adapted to the Jacket of the Army's Individual Protection Outer Garment.

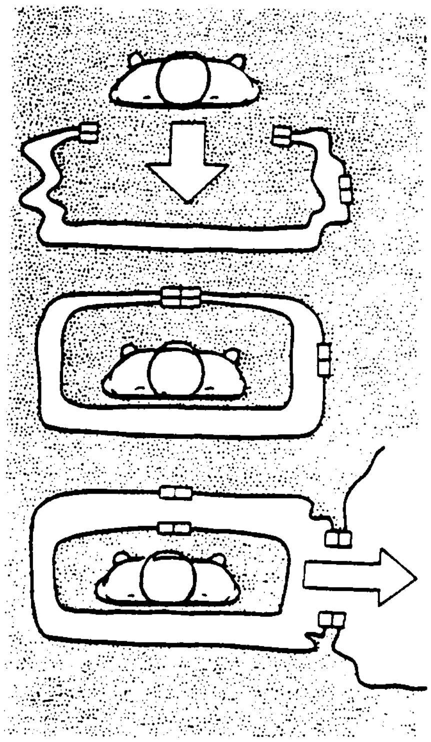


Figure 6- Medical Evacuation Bag. This Application is Designed to Protect Medical Personnel from Chemical Agent Contaminated Casualty During Initial Transport, Treatment, and Decontamination. It consists of Two Bags Bound Together with Interchange Mechanisms so that Space Between the Bags Remains Clean. Casualty is Placed in the Inner Bag and Transported to the Sheltered Medical Center. The Outer Bag is Zipped Onto the Shelter Entrance with a Second Interchange Mechanism and Encapsulated Casualty Moved into the Shelter.

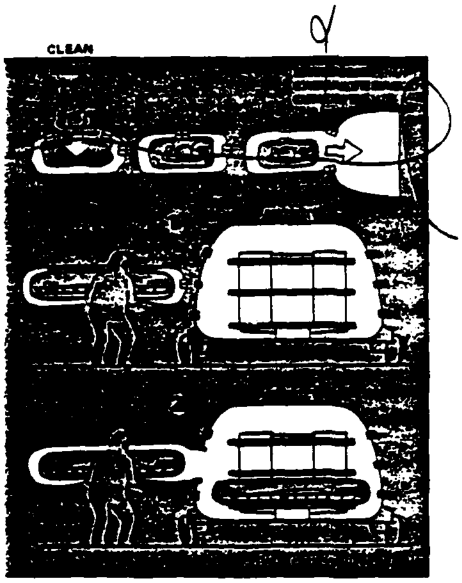
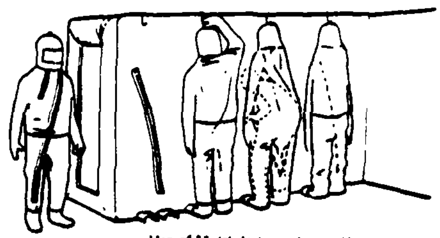


Figure 7- Helicopter Medevac Using Exterior Casualty Bag Concept. A Membrane having Interchange Mechanisms Incorporated in it Permits the Interior of the Helicopter to Remain Clean while Transporting Contaminated Casualties.



Use of Multiple Interchange Mechanisms on the Exterior of Shelters Permits Rapid Entry and Exit of Personnel and Material.

Figure 8- Collective-Protection Rapid Entry/Exit System Sequence of Events

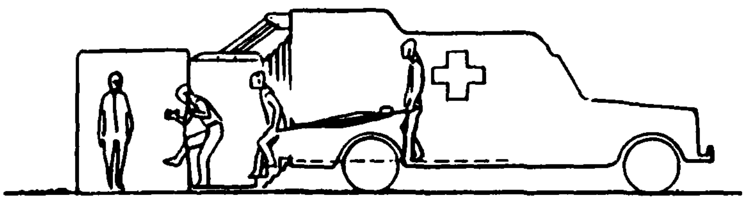


Figure 9- Rapid Transfr. of Personnel and/or Equipment from Vehicle to Collective Protection Shelter Can Be Readily Accomplished With Suitable Entry/Egress System Based on the Interchange Mechanism.

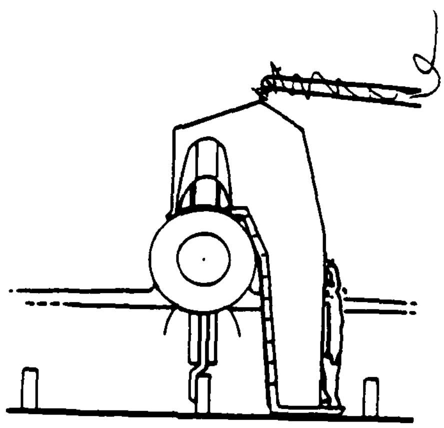


Figure 10- Safe Entrance and Exit of Personnel and Equipment to Aircraft or Land Vehicles Can Be Readily Accomplished with Suitably Designed Systems to be Temporarily Attached to the Vehicles.

PROGRAM PROPOSAL

AN AUTOMATED CLASSROOM FOR ARMOR NCO TRAINING

The Army Research Institute's Fort Knox Field Unit, the Training and Doctrine Command's Training Technology Agency, and the Armor Center are developing an automated classroom for armor NCO training. This training is being programmed for a MicroTICCIT System II, a computer-based training authoring and delivery system. Participants will describe programs that have been written to provide a pre-course diagnostic test on MicroTICCIT, computer-based remedial training over the diagnostic tasks, a complete classroom land navigation training program, free-play surrogate travel training and exercises for land navigation training, computer-based trouble shooting for turret and chassis mechanics at organization and DS/GS levels, and to apply voice recognition and synthesis to armor NCO training tasks in the gunnery and communications areas. The Army Research Institute's program manager for this work will place the effort into the context of an Army drive to create a model technology-based training program and will discuss the problems and pitfalls encountered.

Five papers are included in the program.

Implementation of ATR-Based Courseware on MicroTICCIT: "Navigate from A to B."
Anne W. Martin, Decision and Designs, Inc.

Applications of Voice Technology to Armor NCO Training.
Albert Ingram, James Hassett, and Ruth Ann Marco, Scientific Systems, Inc.

Navigation and Remedial Skills.
Paul J. Sticha and B. Leon Elder, Human Resources Research Organization.

A Model Training Program for Reserve Component Units.
Scott Graham, US Army Research Institute.

The Role of the Training Technology Field Activity.
Donald M. Kristiansen, US Army Research Institute.

Implementation of ATR-Based Courseware on MicroTICCIT: "Navigate from A to B"

Anne W. Martin
Decisions and Designs, Inc.

Abstract

As part of the Training Technology Field Activity (TTFA) at the U.S. Army Research Institute (ARI) Fort Knox Field Unit, Decisions and Designs, Inc. (DDI) implemented courseware to train and test noncommissioned officers on the task of navigating from one point on the ground to another. This courseware is based on the Advanced Terrain Representation (ATR) technology, which provides fully interactive ground-level surrogate travel over open terrain. To perform the work described here, it was necessary to modify a MicroTICCIT workstation to support the ATR technology in addition to designing and implementing the instructional software.

Introduction

The Advanced Terrain Representation (ATR) technology, developed by Decisions and Designs, Inc. (DDI) under the sponsorship of the U.S. Army Research Institute (ARI), is the first implementation of fully interactive ground-level surrogate travel over open terrain. Like earlier "conventional" surrogate travel systems (notably Lippman's ASPEN), ATR simulates travel using images stored on videodisc, accessed through a microcomputer, and presented on a color monitor. Conventional systems let users "travel" down fixed paths, e.g., city streets, by watching fixed sequences of photographs taken along the paths; in these conventional systems, a user may change the direction of travel by selecting another fixed path (sequence) at one of a

---

The work described in this paper was performed under U.S. Army Research Institute Contract No. MDA903-84-C-0333 to Decisions and Designs, Inc. (DDI).

limited number of branch points, e.g., street intersections. ATR also allows users to travel by watching sequences of photographs; however, in ATR, the sequences are not fixed, but are determined by the user. That is, each time a new photograph is presented, the ATR user has the option of changing travel directions, and so determining the next photograph in the travel sequence. Thus ATR lets users simulate travel over open terrain by constructing their own paths through the represented space.

The ideal function for the ATR technology in land navigation training is that of bridge between classroom and real terrain training. The classroom is efficient for the presentation of abstract information and introduction to new principles and skills; real terrain training, by providing fidelity to the situations in which navigation skills are to be used, lends a certain confidence that target skills are in fact being acquired. However, real terrain training is often neither very effective nor very efficient as a use of training resources. Most posts at which training is conducted have only a limited amount of terrain that can be used for training, and this terrain is tightly scheduled; it is difficult to provide students with sufficient time and practice on the training area, and even one day's bad weather can cause significant delays in a class's access to the terrain on which they are to be trained. Instructors are also a limited resource; because students are geographically dispersed during real terrain training, and there are generally fewer instructors than students, student performance often cannot be observed or measured except in a gross, GO/NO GO fashion. Another limited training resource is trainee time. Normally only a limited time period (e.g. three weeks) is available to train students in a whole range of skills, of which land navigation is just a subset. Time spent waiting for the training area to be available, sitting out bad weather, and moving to and from the training area is all "dead time" that could be put to better training use. A final constrained resource in real terrain training is specific to the target population for this pilot application--Armor NCO's. Tanks, along with other tracked vehicles, are very expensive training devices, and their associated logistics (fuel and maintenance) increase the expense and are sometimes scarce themselves. In addition, tanks are destructive to the terrain, and may change training area terrain sufficiently that maps are no longer valid.

The ATR technology can be used to minimize the difficulties associated with real terrain training. The activity that ATR simulates is fundamental to land navigation and many other military tasks--traveling over open

terrain freely. The free travel technology permits one to get lost in the simulated terrain, and the ways of locating oneself in ATR are those that are used on real terrain--identification of terrain features, matching features to a map, and intersection/resection using the features and the map. In fact, most of the navigation tasks performed in land navigation can be performed using ATR; it is expected that the technology can provide a cost-effective replacement for a significant portion of real-terrain land navigation training An ATR-based training device has distinct advantages over real terrain in availability. A device can be used at night and in weather, and there is a certain ease in providing multiples of this relatively low-cost device if more training time is desirable than is available; it is more difficult to increase the supply of pieces of terrain and tanks. Computer-based simulations like ATR can also serve as "force multipliers" for instructors. In addition to presenting training material, the device can monitor student performance and give corrective or reinforcing feedback. Furthermore, ATR provides repeatability of tasks, so that users can apply immediately any corrective feedback or other new information acquired during the training session.

The MicroTICCIT ATR System

Configuration

Figure 1 shows an ATR-capable MicroTICCIT workstation. The computer (an IBM PC), the monitor, and one videodisc player are MicroTICCIT

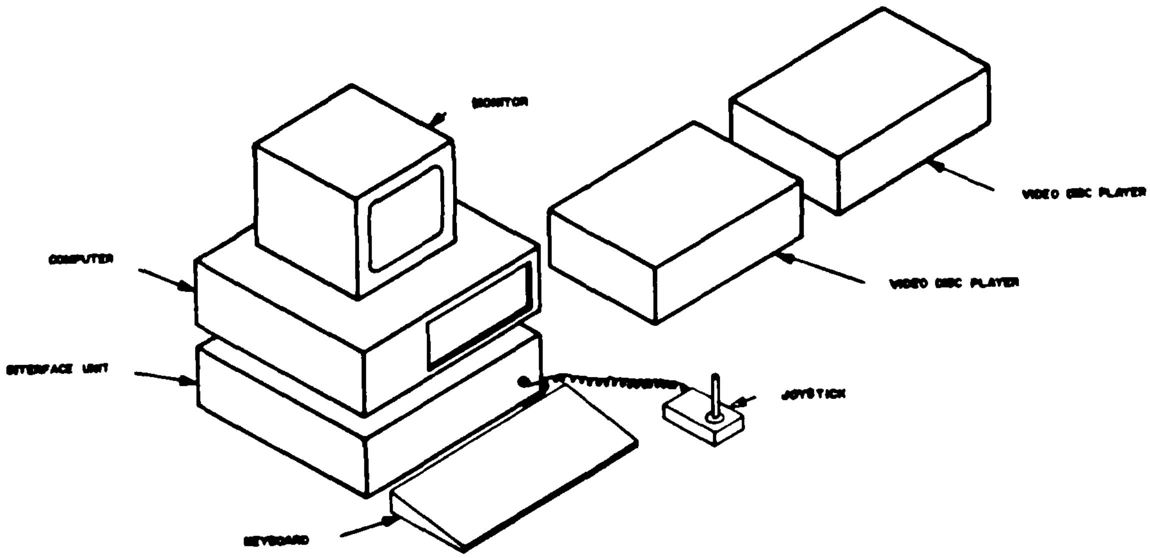


Figure 1. MicroTICCIT ATR System Configuration

equipment; the joystick, the interface unit, and the second videodisc player are additions provided by DDI so that the workstation can run ATR (all normal MicroTICCIT operations can still take place with the ATR modifications in place). The joystick is used by the student to indicate desired direction of simulated travel and to scroll through a menu of training options; it also incorporates a keypad that can be used for communication of numbers (e.g. in the six-digit coordinate system), selection of menu options, and other control functions. The second videodisc player provides continuity of videodisc imagery during travel. The videodisc is used as a random access storage device for photographs. When the user indicates a desired travel direction, and hence the next photograph in the travel sequence, the videodisc player must search for the photograph on the disc, and current player technology does not allow a player to continue sending video signals to the monitor while searching. Thus if only one player were used, the video screen would be blank during search; but when two players are used, one player can display an image while the other is searching for the next image. The interface unit accepts inputs from the joystick, decodes them, and passes them to the computer for processing; it also takes computer output and transmits it to the videodisc players to control search and display, and synchronizes the two players to avoid roll and tear between images.

Courseware for "Navigate from A to B"

The ATR MicroTICCIT land navigation courseware consists of four major functions:

- Free travel--Allows the student to travel freely over the simulated terrain by communicating direction and movement with the joystick. This option gives an oportunity for familiarization with travel system operations before training begins; it may also be used for practicing skills reviewed and learned during other courseware functions.

- Skills review--Presents review material and exercises in prerequisite skills, including the six-digit coordinate system, azimuth determination, azimuth conversion, straight line distance measurement, distance measurement around curves, and map orientation.

- Navigate from A to B--Leads the student through the appropriate

Army lesson plan for navigation from one point on the ground to another, including both a "lecture-demonstrations" and practice problems in which the student's performance is monitored. Both the terrain association and dead reckoning methods of land navigation are explained and exercised.

o Navigation test--Tests students' ability to navigate from one point on the ground to another, using the same testing procedures and pass-fail criteria that are used in administering this test on real terrain. Physical maps and protractors are provided, and the ATR system provides compass information. Three problems are presented, each of which may be solved using terrain association, dead reckoning, or both methods. To begin each problem, students are placed on the simulated terrain and provided with coordinates for the point of placement so that they can locate themselves on their maps. They are then given a destination at which they must arrive using their land navigation skills; they are instructed to travel to the destination and signal when they believe that they have arrived. The criterion for having solved a problem is arriving and signalling within fifty meters of the correct destination, and the criterion for passing a test is solving two out of the three problems correctly.

DISCUSSION

The ATR-based MicroTICCIT land navigation courseware described here is a pilot training application, and is implemented on only one MicroTICCIT workstation. Informally, ATR-based land navigation courseware has been accepted as valid for training by both scientists tasked with providing and evaluating training devices and by military personnel concerned with teaching and using land navigation skills. The ARI Fort Knox Field Unit is currently planning and conducting formal tests of the training validity of the technology. If significant positive transfer of training can be demonstrated, next steps will be to provide more ATR-capable workstations, to develop more videodiscs to extend the simulated terrain base, and to broaden the application of the technology to train more land navigation and other military tasks.

An Automated Classroom for Armor NCO Training: Applications of Voice Technology to Armor NCO Training

by Albert Ingram, James Hassett, and Ruth Ann Marco

Scientific Systems, Inc.

Abstract

One major goal of this project was to identify potential low risk applications of voice technology to Armor NCO training. Given current technical limitations, Scientific Systems recommended that voice technology be employed with computer-based instruction (CBI) only when it is required to meet training objectives, as in training oral fire commands. The second goal was to develop 10 hours of computer-based instruction (CBI) applying voice technology to instruction in fire commands and military communications. The content of this courseware is outlined in this paper, along with the technical problems encountered in interfacing a VOTAN 6050 to a MicroTICCIT CBI system.

Voice Technology and CBI

Voice technology consists of two types of systems that may be relevant to training: voice recognition and voice response. Voice recognition technology is concerned with the recognition of human speech by general purpose computers or dedicated devices. Although large, expensive systems can now accurately recognize techniques to "understand" some speech, the low-cost, low-risk systems that are available are less capable. Such systems must first be trained to recognize a particular speaker's pronunciation of a target vocabulary of 75 to 100 words. 90% accuracy or better is typically expected for recognition of this vocabulary.

The systems tend to be slow, so that fast continuous speech can easily overload them. Unless backed up by large computer systems running artificial intelligence programs they cannot parse the incoming speech or understand it in any meaningful way. Therefore, realistic applications for current low-risk voice technology are limited to recognizing a small number of possible words or phrases and taking an unambiguous action in response to each.

Voice response technology is concerned with the production of speech by computers and other machines. There are several ways of recording voices for later random access and playback, including both analog and digital techniques. The storage media (such as floppy disks, hard disks, and optical disks) for voice playback can hold large amounts of information. A possible limitation is the speed of access. Typically, specific speech segments can be located and played within a few seconds or less. This is fast enough for most situations, but applications requiring extremely fast responses will not be well-served by the technology.

Current voice and CBI technologies are not accurate or fast enough yet to be useful for all types of training. The addition of voice technology to CBI also increases its costs and technical risks. Therefore, training objectives and requirements must be examined carefully to identify areas where voice technology might provide significant advantages over other options.

Scientific Systems therefore recommends that at this time, voice recognition should be used primarily when the training objectives require producing oral commands or communications. Examples are found in areas where commands or communications must be produced orally in actual practice. In contrast, if equal training results can be produced by having the student type, point to, or otherwise indicate his answer silently, then voice technology may not be cost-effective. Thus, voice recognition should generally be integral to the instruction, not merely adjunct to it.

The possible exceptions to this requirement are situations in which the target population consists of people who must enter and receive information by voice. For example, those who cannot read the language of the materials well enough or type and point accurately, may benefit from the use of voice input and output devices. There are also situations in which manual input may be difficult or impossible. For example, if a trainee must handle a piece of machinery in order to learn to maintain it, it might be useful to provide the necessary interactions with voice input and output.

Technical limitations of current systems add two additional constraints: the number of words or short phrases that the voice technology system is required to recognize should be restricted to 75-100 or fewer, and pauses of a few seconds should be acceptable. (Some systems can recognize many different sets of phrases, each one containing 75-100 phrases. Each set must be used separately and independently, so the recognition phrases should be categorizable into distinct groups with 75-100 or fewer elements. Different sets can then be used in different phases of the instruction.)

There are two major criteria for using voice response technology different from the audio playback available from conventional videodisks. First, the situations must require voice or other sound responses as part of a realistic situation or simulation. Second, the requirement for voice production must be separate from any requirements for video and its accompanying audio. This might occur either because video is not required at all or because different audio segments must appear at different times with the video.

Applications of Voice Technology to Armor NCO Training

Scientific Systems applied these criteria in an analysis of the Programs of Instruction for the Basic Non-Commissioned Officers Course, the Advanced Non-Commissioned Officers Course, and the Armor Officers Basic Course to identify specific training tasks that might benefit from the use of this technology.

In general, we found two classes of tasks that are potential candidates. The less promising class consisted of tasks in which it is desirable to have "hands on" experiences with the systems they are learning about. Using voice

technology would allow a trainee to interact with a CBI system and still have his hands and visual attention free to manipulate the materials. For example, erecting a field expedient antenna and completing a radio check using it is a task that could be taught using CBI with verbal instructions, and with a student responding verbally as he completed each sub-task. No demonstrations of the usefulness of this type of training have been made. It might be desirable to do so to help establish its effectiveness, efficiency, and limits.

A more promising group consisted of those tasks which have a high oral content to begin with and that also are good candidates for CBI and interactive video instruction. This class included such tasks as "call for and adjust indirect fire" and "radiotelephone procedures". It also included two tasks which Scientific Systems was contracted to develop courseware for: Fire Commands and Military Communications.

One major task for an M1 tank commander is to direct target engagement activities for his crew. The verbal interactions among crew members to acquire targets, identify them by type and so on have been standardized in a series of verbal fire commands. These commands are currently taught to M1 tank commanders in a classroom environment, demonstrated by an assistant instructor during live range exercises, and practiced on the range. Unfortunately, given existing contraints, there is no time for overlearning these vital commands.

The Fire Commands CBI Unit was designed to teach tank commanders to respond with the correct fire commands quickly and automatically when confronted with a threat. Much of the instruction, and all of the practice, involves the videodisc presentation of filmed targets of various types; the tank commander is given a limited time period in which to speak the correct command for firing upon that particular type of target. Voice response capabilities are used to simulate crew responses in complex interactions. If the tank commander issues the correct command within the proper time, a video overlay shows the target being hit.

Before instruction begins, each student is required to recite the list of command words into the voice recognition device to train it to recognize that particular individual. The first segment of the actual CBI unit introduces the material and provides practice in using the voice recognition device. This is followed by seven segments of instruction and practice that take a learner from extremely simple situations through complex engagements. There are four segments regarding the correct commands for single stationary targets, one for single moving targets, one for multiple targets, and one for simultaneous targets. These are followed by a practice segment including all the types of targets covered previously. Finally, the unit concludes with a formal test of these skills.

The Military Communications CBI Unit is designed to teach M1 tank commanders to maintain the level of reliability and security of tactical communications which will be required for the modern battlefield. More specifically, it teaches the procedures of the Communications-Electronics Operating Instructions (CEOI) required to send coded messages, identify call

signs, authenticate communications, and so on. These procedures are currently taught in a four hour block of classroom instruction. As for the fire commands, there is no opportunity for overlearning.

The Military Communications Unit is divided into two lessons as well as practice and testing portions. The first instructional lesson covers the use of the automated CEOI and related communications materials. It includes an introductory segment, and five instruction and practice segments on finding item numbers, call signs and suffixes, radio frequencies, on encoding and decoding messages, and on security procedures. As for Fire Commands, a pretest is included to help guide the learner into the most appropriate segments. None of these segments require the use of voice technology.

The second instructional lesson covers the procedures to be used on a radio net. It requires the use of voice recognition and response technology to simulate the operation of a radio net. The four segments take the soldier through the processes of signing on and off a net, using standard procedure words, and sending and receiving coded messages. This lesson must be preceded by a separate period in which each student recites the full list of words to be used so that the voice recognition device can learn to recognize his voice.

The instructional segments teach the elements of military communications. They are followed by a practice segment which provides sufficient opportunity for soldiers to practice the complete process of sending messages with standard military communications procedures. Informative feedback is provided to correct learners' errors. The practice unit requires that learners have mastered the skills taught individually in the instructional segments. The posttest segment is similar to the practice segment. However, no feedback is given and complete records are kept of soldier performance.

Technical Issues

Most of the applications discussed here require a voice recognition system that can recognize 50 to 75 words or phrases quickly and with high accuracy. The voice response system must generally be able to produce different voice response segments with random access within a second or two. After a review of available hardware, SSI chose the VTR 6050 voice technology system available from the Votan Corporation to meet these requirements.

At the time of this decision (late 1984), the Votan VTR 6050 was judged the most appropriate for this project. Other candidates which were rejected for this project included the Voice Data Entry System from Keytronics (which lacked a voice response capability), the speech recognition board for the IBM-PC from Tecmar, Inc., of Solon, Ohio (which did not have provisions for speaker-dependent recognition from multiple speakers), and Texas Instruments' Speech Command (which was available only for the Texas Instruments' Professional Computer, and thus could not be interfaced to the MicroTICCIT system). It should be noted that these technologies are changing quite rapidly, and that a project begun at this time would require a new survey to identify appropriate voice systems.

A number of technical problems were involved in interfacing the VOTAN device to the MicroTICCIT system in use at Fort Knox. This system consists of a Data General Eclipse host computer which is connected to a cluster of modified IBM-PCs which serve as workstations. Each MicroTICCIT workstation which is to be used for voice recognition must be able to send commands to the VOTAN (e.g., "now listen for a word") and receive responses from the VOTAN (e.g., "the word 'gunner' was spoken"). This requires both hardware and software modifications to the current system.

The hardware modification is quite straightforward: an RS232 asynchronous communications card must be inserted into one of the expansion slots of each TICCIT workstation which will be used to deliver this courseware. Unfortunately, the software modification proved to be far more complex.

Each MicroTICCIT workstation is actually an IBM-PC with a special keyboard and several special cards. It can operate in two different modes: as a workstation on a MicroTICCIT network, or as a stand-alone IBM-PC. For the stand-alone IBM-PC mode, the program which controls communications with the VOTAN is relatively straightforward. SSI wrote this program in C, and the project will use this mode of operation for voice recognition training of the VOTAN.

The problem of contolling commmunications with the workstation while it is functioning on the MicroTICCIT network is far more difficult. An IBM-PC assembly language program was required to control communications between the IBM-PC and the VOTAN under TPOS, the system developer's Terminal Processor Operating System for the IBM-PC. Since this operating system is unique to TICCIT and little documentation was available, SSI spent several weeks consulting with the developer's software engineers.

However, a workstation in the MicroTICCIT network is ordinarily controlled by a program which is written in the ADAPT authoring language and which resides on the Data General Eclipse. Thus, the ADAPT program which runs all of the courseware for this project must be able to send commands to SSI's assembly language communications program, and receive information from it. The ability to access independent programs written in other languages from within an ADAPT program was developed by the TICCIT developer's software engineers for this project. A second capability was also added to the ADAPT language for this project: another new ADAPT command was required to download the assembly language program for controlling the VOTAN from the Data General minicomputer to each workstation that has a VOTAN.

COMPUTER-BASED TRAINING FOR M1 TANK COMMANDERS: NAVIGATION AND REMEDIAL SKILLS

Paul J. Sticha
B. Leon Elder
Human Resources Research Organization

Abstract

The Training Technology Field Activity (TTFA) at Ft. Knox has been established to apply recent developments in technology to meet the training needs of M1 tank commanders. To support the TTFA, computer-based instruction (CBI) has been developed to teach navigation skills, and to provide remedial training in prerequisite skills. The instruction applies recent advancements in video technology using a state-of-the-art CBI authoring and delivery system. This paper discusses the tasks that are trained, the technological alternatives considered, and describes some of the methods used to present the instruction.

Introduction

The U. S. Army established the Training Technology Field Activities (TTFA) in response to the recognized need to improve training effectiveness by appropriate application of recent developments in training technology. Of particular concern to the TTFA was the application of computer-based instruction (CBI), hand-held computers, computer speech technology, and videodisc to identified training needs. In supporting the TTFA at Ft. Knox, the Human Resources Research Organization (HumRRO) has developed CBI for the 19K Basic Noncommissioned Officer's Course (BNCOC). The instruction addresses skills in land navigation, and provides diagnostic testing and remedial training for selected prerequisite skills. The instruction relies heavily on the use of video, which is presented using the MicroTICCIT system.

This paper describes the results of our efforts, and gives some of the rationale behind the choice of presentation media for instructional elements. The first section describes the tasks for which training was developed. The second section describes the factors considered to determine the appropriate presentation media for instructional elements. The third section describes the structure and content of the resulting instruction.

Training Tasks

The instruction that HumRRO produced comprises two of the eleven modules of the 19K BNCOC course. The first module covers diagnostic testing and remedial instruction for selected prerequisite tasks; the second module provides training in navigational skills.

Prerequisite Tasks

The prerequisite tasks selected for diagnostic testing and remedial training were identified in earlier analyses of the 19K BNCOC (Drucker, Hannaman, Melching, & O'Brien, 1984; Drucker & O'Brien, 1985). Five tasks were chosen from 21 candidates because of their judged suitability to CBI:

1. Determine grid coordinates of a point on a military map using the military grid system.
2. Communicate using visual signalling techniques.
3. Recognize and identify friendly and threat armored vehicles.
4. Establish tank firing positions.
5. Operate radio set.

Available documentation on these tasks was gathered, and the tasks were analyzed to determine task elements, conditions, and standards. In addition, performance data were collected to identify common errors, and to assess mental structures that experienced soldiers use to organize the information required to perform the tasks (Knerr, Elder, Campbell, Harris, Stein, Sticha, Morrison, & Russo 1985). The results of these analyses were used to guide the training design.

Navigation Tasks

The ARI Field Unit at Ft. Knox convened a Land Navigation Steering Committee which identified ten tasks as candidates for CBI. From these tasks, five tasks were selected to develop training according to the judged training need, and the judged efficacy of CBI compared to alternative training methods, such as classroom instruction and field exercises. The following tasks were selected for training (see Knerr, Sticha, Elder, Ramsberger, Harris, & Tkacz, 1984):

1. Analyze terrain using the five military aspects of terrain.
2. Identify natural terrain features and determine elevation.
3. Orient a map to ground by map-terrain association.
4. Determine location on the ground by terrain association.
5. Locate an unknown point on a map or on the ground by intersection or resection.

Because a compass is not accurate in close proximity to a tank, the tank commander must rely on his ability to relate terrain with its associated map representation in order to determine his location, or the location of other points. Consequently, the instruction stresses methods of terrain association in its instructional approach to all tasks.

Selection of Presentation Format

The system chosen to implement the instruction (MicroTICCIT) integrates several presentation media, including text, video, graphics, and audio. Part of the process of instructional design involved the selection of appropriate media to present instructional segments. The selection was based on task characteristics as well as the capabilities of the instructional system.

Video was chosen to present all instruction in the navigation tasks, and selected prerequisite tasks. Instruction, in these tasks, consists of a live-action demonstration, which is ideally suited for video. Video is also used to present still pictures of terrain, maps, and armored vehicles. Finally, video is used as the basis for simulation in two navigation tasks.

Graphics is used to enhance information presented on video. Graphics overlays are used to highlight areas of the video presentation, organize information, and indicate response areas. In addition, graphics are used as the basis of a simulation for the task, Operate the Radio, where video simulation proved infeasible.

Text is used to present instruction on operating the CBI system. Since details of the user-computer interaction were likely to change from the initial design, it was not considered wise to present instructions on operating of the system on a medium that did not permit easy editing. In addition, text and audio are often used to reinforce each other. That is, instructions that appear on the terminal screen in text form are often read to the trainee over the audio system.

These decisions on the use of media depend on the particular capabilities of the instructional delivery system, and would have been different with a system with different capabilities. For example, MicroTICCIT has the ability to display high resolution graphics quickly. This capability makes graphics a good choice as the basis for the simulation in the radio task, in which a video simulation would be cumbersome, because of the time required for videodisc searches. In a system using a more sophisticated videodisc, however, video might have been a viable alternative for this task.

Structure and Methods of the Instruction

To enhance the clarity of the instruction, we provided a common structure of instructional segments. The structure defines instructional components, methods for displaying menus, and uses for color. The methods of instruction, on the other hand, are tailored to the requirements of the training tasks. We obtained guidance for instructional design from the Training Effectiveness, Cost Effectiveness Prediction (TECEP) methods (Aagard & Braby, 1976).

Instructional Components

With minor exceptions, instruction for each task consists of seven components: objective, instruction, review, help, practice, practice help, and test. All components except the objective and practice help are accessible from a single menu presented after the objective. The objective, which presents a brief statement of what the student is expected to do after receiving the instruction, is automatically displayed whenever the student enters a lesson. Practice help is only accessible from the practice problems.

Three components present the material to be learned in slightly different ways. The instruction section provides a complete exposition of the required material, along with examples and opportunity for student interaction, when required. The review section provides a brief summary of the instruction, outlining the procedural steps, or summarizing concepts. The instructional help section provides instruction on prerequisite skills, corrects common errors, or provides a simpler presentation of the material.

The practice section provides the student an opportunity to practice procedures, or answer questions about the material in the lesson. Problem-specific help is available in the practice section; in some lessons the student may select the help, while in other lessons help appears automatically whenever the student makes an error. In general, the lesson tests are the same as the practice problems, except that tests are graded while practice problems are not.

Instructional Material

The instructional methods were determined by task requirements as adjusted by the capabilities of the instructional delivery system. The capabilities of MicroTICCIT allow several interactive methods to be incorporated into the instructional materials. For example, some exercises for navigation tasks allow the student to scan the horizon, locate prominent terrain features, and determine his location or the direction he is facing. This simulation is accomplished by showing the student a sequence of still photographs stored on videodisc. The still photographs are arranged in such a way so that when they are shown at the proper rate, they give the student the impression that he is turning. The CBI system keeps track of the direction the student is looking by knowing the frame that is being displayed to the student at any time.

Another simulation involves the operation of the radio. In this simulation, the student is presented with a graphical picture of the radio. When the student touches appropriate switches, they change position. The student may set the position of the various switches until they are in the correct position.

Conclusion

The instruction developed for the 19K BNCOC course illustrates ways technological advancements in video and CBI can be applied to meet the training needs of the tank commander. Future technological advancements will further enhance the effectiveness of CBI. These advances, including decreased video search time, single-frame audio storage, and increased graphics capability, will increase the range of tasks that can be addressed effectively by CBI. However, even current capabilities can support effective training for a variety of skills.

Acknowledgements

The author would like to thank the many individuals who contributed to the effort reported here, including Mazie Knerr, Roy Campbell, Peter Ramsberger, Carolyn Harris, Doris Stein, John Morrison, and David Hopwood. The project was sponsored by the Defense Advanced Research Project Agency, the U.S. Army Research Institute, and The Adjutant General's Office, under Contract Numbers MDA903-83-C-0453 and MDA903-84-C-0479. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U. S. Government.

References

Aagard, J. A., & Braby, R. (1976). Learning guidelines and algorighms for twelve types of training objectives (TAEG Report No. 23). Orlando, FL: Department of the Navy, Training Analysis and Evaluation Group.

Drucker, E. H., Hannaman, D. L., Melching, W. H., & O'Brien, R. E. (1984). Analysis of training requirements for the Basic Noncommissioned Officer Course for M1 Tank Commanders (19K BNCOC). Alexandria, VA: U.S. Army Research Institute.

Drucker, E. H., & O'Brien, R. E. (1985). The development of training and evaluation products for the basic noncommissioned officer course for M1 tank commanders (19K BNCOC). Alexandria, VA: U.S. Army Research Institute.

Knerr, C. M., Elder, B. L., Campbell, R., Harris, C. D., Stein, D. J., Sticha, P. J., Morrison, J. E., & Russo, R. P. (1985). Training analysis and design for remedial computer-assisted instruction for tank commanders. Alexandria, VA: Human Resources Research Organization.

Knerr, C. M., Sticha, P. J., Elder, B. L, Ramsberger, P. F., Harris, C. D., & Tkacz, S. (1984). Interactive videodisc training for army land navigation skills. Alexandria, VA: Human Resources Research Organization.

An Automated Classroom for Armor NCO Training: A Model Training Program for Reserve Component Units

Scott E. Graham
U.S. Army Research Institute
Fort Knox Field Unit

Abstract

The Model Training Program for Reserve Component Units (MTP-RC) has developed and evaluated 200 hours of simulation courseware for training M1 tank maintenance skills. The paper describes the top-down courseware development process used to produce generic instructional templates which lead to enhanced production efficiency and improved communication between instructional designers. The primary goal of the courseware is to train soldiers to follow the troubleshooting procedures in the M1 Technical Manuals. The courseware description demonstrates how a lightpen is used to move around in the simulated tank, connect test equipment, and receive diagnostic readouts.

Introduction

Maintenance units in the Reserve Components (RC) are facing significant training challenges as new high technology weapon systems are introduced into the Army inventory. Some National Guard combat units are being issued the same M1 tanks, Bradley Fighting Vehicles, and M60A3 tanks that active forces are receiving. RC Combat Service Support units have post mobilization responsibilities that will require them to maintain new equipment. The challenge will be for RC maintainers to attain and sustain acceptable levels of proficiency on new weapon systems under adverse training conditions.

The Model Training Program for Reserve Component Units (MTP-RC) is a joint project of the Army Research Institute and the Training and Doctrine Command's Training Technology Agency which is attempting to produce at least a partial solution to the RC maintenance training problems. RC units are typically limited to 39 training days per year, split between a 2-week summer camp and monthly drills. In addition, these units are frequently short of trained cadre and training support materials. For example, a reserve maintenance battalion which under the CAPSTONE Program would round out a division with M1 tanks has little access to these tanks for training. This situation is therefore appropriate for simulated M1 maintenance training. Also, the distributed nature of the monthly weekend training is well suited for a CBI system which can monitor individual progress and direct training.

Computer-based instruction has the potential to address many of the specific problems found in maintenance training. Reduced training time which typically results from CBI could allow more maintenance issues to be covered within existing time constraints. Soldiers could also be trained to standardized levels of performance across various units. CBI can teach theory, in addition to procedural steps, and the instruction can readily be paired with simulated hand-on experience. Lessons on troubleshooting can require the student to carefully follow procedural steps found in TMs, and at the

same time explain why the steps are being performed. Lastly, CBI reduces the reliance on Actual Equipment Trainers (AET) which can be expensive, dangerous to work on, and in the case of the M1 tank, scarce.

Approximately 200 hours of MicroTICCIT courseware for training four Military Occupational Specialties (MOS) is being developed under contract. The courseware primarily trains skill level 2 maintenance tasks for turret and hull mechanics at the organizational and direct support/general support (DS/-GS) level. The primary goal of the MTP-RC is to sustain RC maintenance skills on equipment systems which are not physically available. This is accomplished by ensuring that soldiers have the fundamental skills to use the Technical Manuals (TM) by including both remedial instruction and repeated practice in using the TM to troubleshoot simulated equipment.

MTP-RC Courseware Development Process

The MTP-RC project developed a large amount of courseware in a limited amount of time. To accomplish this, the courseware development process made several refinements on the traditional Instructional Systems Design (ISD) model which have lead to greater production efficiency and better communication between instructional designers. The process used a top-down, macro design approach which focused on commonalities within the content domain and resulted in the development of instructional templates which could flexibly be used over and over again.

The MTP-RC process began with the Ordnance School selecting tasks with "High Field Utility", as a function of criticality and the liklihood that the diagnostic procedures would not soon change. In addition, Subject Matter Experts (SME) selected maintenance tasks which involved Line Replacable Units (LRU) with higher estimated failure rates. In an attempt to maximize the transfer and generalization of the training, tasks were also selected to provide experience with as many LRUs as possible.

Traditional maintenance training typically focuses on successful whole-task completion, as the soldier is trained by following the TM from the beginning of a procedure to the end. The macro design approach differs from the traditional by partitioning the whole-task into functional subtasks. The focus then falls on the cognitive and psychomotor "enabling" skills necessary for successful subtask performance. The cognitive enabling skills identified primarily involved knowledge of system operations and the effects of various malfunctions on system performance. The psychomotor enabling skills were clustered as either check/inspect or maintenance actions. The check/inspect tasks involved visual inspections, electrical tests, and mechanical tests for determining faulty components, while maintenance actions required replacing parts, lubricating, and using special lifting devices. Each of the cognitive and psychomotor enabling skills defined specific training needs which subsequently were included in one or more lesson specifications.

Broad "Generic Enabling Objectives" were then formed by grouping the enabling skills by commonalities. These generic enabling objectives were quite general such that several entire lesson segments would fall under one. For example, "Describe the normal operation of a tank system" would be superordinate to all tank systems trained. From these generic enabling

objectives, the five basic templates were derived from which all 200 hours of courseware would be developed.

Decisions were then made for each of the templates as to the instructional strategies to be used and the nature of the student interactions. In defining the instructional components, for example, it was decided that the Name-Locate-Function (NLF) template would contain five sections, namely an introduction, overview, instruction, practice, and review. At a lower level, the introduction would contain four components: a brief description of the lesson topic, a statement of objectives, a description of the importance of a lesson with any unusual features it might contain, and a description of a relationship between that lesson to other lessons. These instructional components describe the type of content which will be added later, and not the content per se.

The template also defines how the learner will interact with the content and how the system will respond to those interactions. Of particular interest are the "recurring interactions" which were used repeatedly. The pseudocode which is written to specify the template, for example, might define a subroutine, "GETINPUT", which takes the soldiers lightpen response from the screen. Once established, the template pseudocode must only specify "Do GETINPUT" each time this action is required. A critical requirement of the pseudocode is that it contains sufficient detail for programmers to use as the basis of the program code, yet remains simple enough for other instructional designers to know what is happening.

The templates were written by the senior instructional designers and underwent a two phased review. On the technical side, a senior programmer evaluated whether the template plans were within the limits of the authoring and delivery system, the type and amount of branching required, the amount of information presented per page, and estimated the personnel requirements for writing the code. The programmer suggested, when appropriate, how alternative strategies might be more efficient. A second instructional designer also reviewed the templates for possible improvements in the generic instructional approaches. Following these reviews, ADAPT programming code was written and debugged for a prototype lesson. While the template development required a significant amount of time early in the project, SSI reports the templates resulted in an overall time savings of around 25%. In addition, the instructional templates can be used in later CBI projects.

A key to the potential success of the program was the pairing of professional Instructional Designers (IDs) with SMEs in the development process. Based on the templates and Lesson Specification Documents, the SME/ID team wrote production specifications for the courseware. These specified what text and graphic or video displays were needed for each courseware segment. As a means of quality control, a second SME/ID team reviewed each production specification. Actual courseware production was performed by a Courseware Developer, who wrote the ADAPT code and input the text, and a Graphics Specialist. Following initial debugging, the courseware was reviewed by both the original and second SME/ID teams, and send back to the Courseware Developer for corrections.

The segment was then sent to the Ordnance School were it was reviewed for technical accuracy by SMEs and instructional approach by an Education

Specialist. Special attention was paid to details of dress and SOPs, for which the military user is a stickler. A formative evaluation was then conducted with pilot soldiers to identify any problems not surfaced in the development reviews, e.g., inappropriate reading level of the text. Additional courseware modifications were made prior to the final delivery, and these primarily involved the updating of TM references.

The iterative review process, while costly and time consuming, may be necessary to guarantee quality in the final product. There is currently a push in the CBI industry to develop courseware authoring systems which can be used with little instructional design or computer experience. The training value of courseware resulting from these "easy-to-use" approaches is, however, suspect. This is not to say that courseware authoring systems are not expected to become easier to use over time, but that training of complex skills requires appropriate expertise.

Courseware Description

The introductory course begins with a set of lessons which trains the soldier how to use the MTP-RC training program on MicroTICCIT. The courseware is designed such that once the soldier is logged onto the system, all interactions are done with a light pen, i.e., without the keyboard. The student can then receive refresher training on using the Simpson Digital Multimeter, Breakout Box, and STE-M1 test sets. Also included are reviews of how to use the technical manuals with practice exercises and tank safety requirements.

The remaining four MOS specific courses are organized by tank systems. Each unit begins with a lesson that describes the principles of a particular system, e.g., the fuel supply system. The first segment describes the Name, Location, and Function of each part within that system. A second segment, built from a different template, discusses the Input, Process, and Output of these same components. This basic structure is repeated in each Principles of Operation lessons. Given the RC weekend training schedule, the uniform template structure should help students remain familiar with the lesson structure from month to month and should result in reduced training time.

The primary objective of the courseware is to train soldiers to use the TMs to troubleshoot simulated M1 tank systems. To this end, most of the MTP-RC courseware consists of troubleshooting segments. Each troubleshooting lesson begins by introducing a particular symptom within the system being trained. The introduction includes a conceptual explanation of what system components are possibly causing the fault. The troubleshooting lesson then presents a "Guided Demonstration" for troubleshooting that symptom in which each procedural step from the TM is cued on the screen. Two "Practical Exercises" follow for the same symptom, each terminating in a different fault. For example for Fuel Supply System-Fault #5, "Fuel gage reads zero in all fuel tank selector switch positions", one exercise branches to find a faulty Hull Networks Box while another identifies the Driver's Instrument Panel.

The courseware requires the student to read the TM while troubleshooting. By using the lightpen, the soldier interacts with high resolution color graphics and is able to move around in the tank, connect simulated test equipment, and receive diagnostic readouts. Action "icons" or graphics at

the bottom of the screen permit the student to connect, disconnect, inspect, remove, or replace parts and equipment.

If the student reads, "Connect red multimeter lead to point 16 on the breakout box," the student would first touch the "connect" icon and then the red lead. The screen would show, "Connect red lead to what?". The student would touch the appropriate point on the breakout box and the graphic would change to show the connection had been made. Feedback is given after each step, and when an error is made, the correct step is identified with a green graphic overlay.

The 63H courseware for DS/GS hull mechanics trains the identification and repairing of bad engine and transmission parts. The majority of the interactions in the 63H maintenance simulations are with videodisc pictures of actual equipment rather than computer graphics. The soldier must, for example, remove brake pads or gear sprockets, inspect them for wear, and then replace the parts.

At any point during the troubleshooting, various types of advice are available. The soldier can get an explanation of why the current steps are being performed, and frequently a detailed wiring diagram. Other types of advice include descriptions of the icons and information on the correct TM page number and next troubleshooting step.

Program Evaluation

A series of transfer evaluations conducted at the Ordnance School, APG, and the Armor School, Ft Knox, has demonstrated the courseware's training effectiveness. Soldiers who received the simulated troubleshooting training made fewer errors on hands-on troubleshooting tasks than did soldiers trained under conventional methods. The skills and knowledge developed in the exercises also generalized to troubleshooting a task not specifically trained. This generalization was likely the result of soldiers being trained and given practice on properly using the TM.

The value of this program, and potentially others like it, will be better understood when the training systems are placed in reserve units. Three RC units have been selected for trial implementation in mid-FY 86. They are the 195th Maintenance Company, Westminster, MD; the 2198th Maintenance Company, Dagsboro, DE; and the 2/522d Armor Battalion in Raeford, NC. Among the important questions addressed in the project are whether FORSCOM will consider this form of CBI simulation a valuable training tool, and whether the units can and will use the training. If so, courseware development, hardware, and system maintenance costs need to be built into training budgets.

Despite limited training time and resources, RC units are required to develop and sustain wartime proficiency on high-technology weapon systems which may not be present. No one approach can obviously overcome all of the deficiencies in RC maintenance training. Computer-based training approaches, such as the Model Maintenance Training Program for RC Units do, however, provide another resource for training maintenance skills and knowledge. The training is exportable, requires no additional instructors, and does not rely on the presence of an M1 tank. As a complement to present RC programs, the Model Maintenance Training Program can enhance RC proficiency.

An Automated Classroom for Armor NCO Training: Role of the Training Technology Field Activity

Donald M. Kristiansen

US Army Research Institute for the
Behavioral and Social Sciences
Fort Knox, Kentucky

Abstract

The US Army Research Institute and the US Army Training and Doctrine Command have established a jointly staffed Training Technology Field Activity (TTFA) at Fort Knox, KY to identify, develop, implement, and evaluate a variety of new, low-risk, technology-based training methods and programs. Many of these new programs are being developed using computer-based instruction (CBI) in a direct effort to address the need for higher quality training which is coupled with a steady reduction in training resources, particularly instructors. As part of this effort, the TTFA at Fort Knox is developing an automated classroom for training M1 Tank Commanders. This paper describes the effort to develop an automated classroom and discusses the unforeseen problems encountered.

Background

While the Army has adopted many modern training technologies, the most common training delivery system is still the stand-up lecture. This delivery medium is becoming harder and harder to sustain, however, as instructors, like other resources for training, are becoming scarce. To overcome this trend, and to generally increase the quality of instruction, the Department of the Army has directed research to identify and implement promising new methods and approaches to training. Historically, this research has not yielded a big payoff.

In an operational training environment it is hard to test and apply promising technologies, especially given the rapid rate at which technologies are being developed today. Courses have fixed lengths and are tightly scheduled, resources are in critically short supply, course managers have all they can do to see to it that scheduled activities take place in a manner close to what was planned, and instructors view change critically and often resist it. In a training development environment it is hard to determine the key training problems and to ensure that useful results of technology development are appropriately applied. Given the complexity and cost of new training systems, there is a need to establish an environment in which technology developers and trainers can work together continuously to test and apply training innovations in a systematic fashion.

This need has led the US Army Training and Doctrine Command, the US Army Armor Center, and the US Army Research Institute to develop a cooperative working partnership operationalized as the Fort Knox Training Technology Field Activity (TTFA). The mission of the TTFA is to identify, develop, implement, and evaluate a variety of new, low-risk, technology-based training methods and programs. Technology has been defined to include techniques, strategies, methods, and models as well as hardware, software, and courseware. The initial Fort Knox test bed for introducing training innovations is the M1 Tank Commander Course (19K BNCOC).

CONCEPT OF AN AUTOMATED CLASSROOM FOR ARMOR NCO TRAINING

Analysis of the 19K BNCOC pointed to several training problems that ran across the various blocks of instruction. Instructor resources were critically thin. In many instances practice activities for skill development were nonexistent because there were not enough instructors to monitor practice and provide the necessary feedback. This problem did not appear correctable. In fact, it will probably get worse as more and more slots are transferred from the training base to the cutting edge of the Army, operational combat units. A second problem was inconsistent instruction. To a considerable extent training was instructor dependent. What was taught, how it was taught, and the quality of instruction varied greatly across instructors. NCOs were advancing to higher level instruction without having mastered prerequisite instruction. This was, partially, a resource problem (time and instructors) and partially a lack of quality control. A reasonable approach to avoiding all these problems, and, at the same time, generally raising the quality of instruction, was to develop computer-based training (CBT) and testing for all feasible tasks in the course.

The delivery system selected for CBT development was the MicroTICCIT System II with a videodisc overlay option. Each MicroTICCIT workstation is a multipurpose station designed for use by students, instructors, courseware developers, and training administrators. It consists of an IBM personal computer, a Sony color monitor, a laser videodisc player, a detached keyboard with an integral instructional keypad and special function keys, a light pen, and a high speed communications link to the system host micro-computer. The display supports flicker-free presentations of computer-generated text and graphics and is fully compatible with videotape and videodisc. Under author control, up to seven foreground colors and four background colors can be selected for use in the creation of each frame of computer-generated text and graphics. The colors are selected from a palette of 4,096 colors. The videodisc overlay option allows simultaneous display of computer-generated text and graphics and NTSC video combined on a single MicroTICCIT display. The videotape image, which can include either still frame or motion sequences, can be presented on all or any part of the display screen and can be overlayed by computer-generated text or graphics to single pixel resolution.

To maximize the potential of this system and meet the particular training needs of the 19K BNCOC, a series of CBT development contracts were let. These contracts provided MicroTICCIT software and courseware to computer manage the course, a pre-course diagnostic test and computer-based remedial training over the diagnostic tasks, a complete classroom land navigation

training program, free-play surrogate travel training and exercises for land navigation training, computer-based trouble shooting training for turret and chassis mechanics at organization and DS/GS levels, and an application of voice recognition and synthesis to Armor NCO training tasks in the gunnery and communications areas. Each of these applications is addressed in this symposium.

PROBLEMS AND PITFALLS

Self-Inflicted Wounds

Several of the courseware development efforts included an attempt to stretch the capabilities of the MicroTICCIT beyond its original design. Particularly troublesome were the interfacing of the voice recognition and synthesis system and the surrogate travel system with the MicroTICCIT. A support contract should have been let with the MicroTICCIT developer so that they could have worked with the TTFA contractors to modify the MicroTICCIT for these applications. The TTFA contractors were continually in need of support from the hardware developer. The MicroTICCIT developer did provide considerable support in the absence of any remuneration but this support necessarily took place when time could be made available from ongoing projects. These TTFA applications eventually required the developer to add to a software update. A separate contract with the MicroTICCIT developer would have provided timely support that was focused on the particular application in need of their services. As a result, some of the software is not as efficient as it could be.

Some things that should have been patently obvious just slipped through the cracks. For example, no one thought to check on the printing capabilities of the MicroTICCIT System II. The contractor developing the course management package quickly found out that the hardware provided by the government did not have a printing capability. This presented a severe constraint to the efficient debugging of code. Code had to be brought up on the screen and then copied by hand to provide a listing.

The early decision to develop courseware on the MicroTICCIT was made at a time when the System II was just beginning to be marketed. A host and three workstations were made available at Fort Knox out of another research effort which was also using this system. In addition, a host and several workstations were made available to courseware developers in the Washington, DC area. The early notion was to use this equipment to determine if useful courseware could be developed on the MicroTICCIT and to conduct the necessary training and transfer trials. Little thought was given to what happens next, namely, implementation in 19K BNCOC. Four workstations are not adequate to field all the developed courseware in the course. 19K BNCOC is expected to grow in size by a factor of seven by FY87. The Armor Center has adopted the reasonable position that courseware will not be implemented until enough hardware and courseware are on-hand to preclude scheduling nightmares. We now find ourselves in the position of having courseware but with too few workstations to implement. It is possible that the courseware, and, perhaps

the notions of CBT and TTFA, will be rejected, not because the courseware,or CBT, or TTFA, are lacking, but rather because there was not enough hardware to play the courseware for the majority of the course at one time.

Acquiring the necessary hardware on a crash basis (under 12 to 18 months) has proven to be almost impossible. The approval chain for the purchase of workstations goes all the way to the Secretary of the Army level.

Unanticipated Problems from the Outside

No one staff agency had responsibility for identifying site requirements and then matching those requirements with available (or developed) facilities. No SOP existed for this. the Directorate of Information Management was interested in tracking our progress and was willing to help but was clearly not charged with any responsibility in the matter. TTFA personnel forged a path here through some pretty thorny thickets. After finding three suitable sites, which all belonged to someone else of course, the TTFA had to find out which agencies had to be coordinated with to ensure approval. A site was finally selected by fiat by the Deputy Assistant Commandant. Coordination became necessary with the Fire Marshal because automatic sprinklers had to be turned off and a halon system installed; with the Provost Marshal for physical security requirements; with the Director of Information Management for compliance with regulations and SOPs governing computers and computer networks; with the Director of Security for potential security leaks and/or violations; and with the Engineers for site preparation. The number of electrical outlets had to be increased fourfold. Extra air conditioning had to be installed. A power management system was needed. Secure doorways had to be installed. Initial site preparation costs exceeded $20,000 in a room that already was physically secure, had some air conditioning, and was already attached to a building-wide power management system. As a result of all this, the MicroTICCIT had to be set up in a temporary site for months. While in the temporary site it was almost destroyed by a lightning strike. Nothing in any literature, or in any conversations with colleagues, forewarned us of this. We were unprepared for this problem.

The TTFA partnership, established and signed at the General Officer level, has created unforeseen problems at the staff agency level. If the TTFA is successful in developing and implementing a high-tech BNCOC, it will be a unique BNCOC since the TTFA is changing what is taught as well as how it is taught. The other two existing 19K BNCOCs do not belong to the Training and Doctrine Command. One belongs to the Forces Command and the other belongs to the Seventh Army in Europe. Hard planning for the salesmanship effort that will eventually have to be launched to get these major commands to change their BNCOCs has not started. Serious planning will undoubtedly await a "success" story at Fort Knox. That will mean that implementation in the other major commands will not follow for at least two years after a successful demonstration at Fort Knox. Meanwhile, the Soldier's Manual, extension training materials, and the Soldier's Qualification Test will be based on the BNCOC currently on-going. Voices are already reminding us that that will put the Fort Knox BNCOC graduate at a disadvantage.

Traditional instruction, that it, the methods and procedures in current vogue, is embedded in an Army School System that has been designed and staffed to receive and modify it. The mechanisms by which instruction is developed, staffed, approved, resourced, and implemented are in place and are seen as "the way it has to be done." No such mechanisms are in place for automated instruction. CAI is on-going at many installations but we could not find anyone at Fort Knox that knew of established procedures for instituting such a system. All the traditional mileposts for instructional development, the mileposts that the staff agencies look for and guide on, are missing when you introduce automated instruction. Lesson plans are different, course designs differ radically from traditional course designs, course design and development was done by contract agencies because neither ARI nor the Armor Center had in-house resources to do this, the physical document that is processed is a piece of computer software rather than paper copy, etc. Since the Army, like all military institutions, is rule-bound, and there were no rules readily apparent for the introduction of automated instruction, no one in the official chain for staffing changes or additions to instruction knew how to handle these materials.

Directorates have been established under the Army School Model to process the flow of materials and to administer the decision-making process according to current regulations and SOPs. These regulations and SOPs call for certain products to arise from the instructional development process. Formats are specified for these products. Staff agencies are charged with ensuring that instructional materials such as task statements, training design documents, staffing and approval documents, and lesson plans follow the correct formats and the correct processing procedures. The instructional design process and the resulting formats that we followed under TTFA did not conform to these regulations and SOPs. As a result, no one quite knows how to make the Micro-TICCIT instruction an official part of the course.

Despite these problems, TTFA expects to have a demonstration "automated classroom" in place at Fort Knox by mid-FY87.

The Measurement of "Soldier Will" and Its Relationship to Well-Being, Life and Army Satisfaction, Duty Stress, Health Problems, and Unit Replacement

James Griffith, Ph.D.
Walter Reed Army Institute of Research
Washington, DC 20307-5100

Abstract

Aspects of the the Army's New Manning System (NMS), such as group training, transfer of personnel in groups, and low personnel turbulence within the unit during the initial first-term enlistment were to build strong interpersonal relationships among soldiers and between soldiers and officers. These changes, in turn, were expected to have both individual and group effects pertinent to the soldier's will to fight. A crucial element in evaluating whether NMS has these intended human effects is defining the psychological phenomena that compose the soldier's psychological readiness to fight (collectively known as "soldier will"). The present study developed reliable, valid, and useful measures of "soldier will." Questionnaire responses given to morale and cohesion scales by 2830 soldiers (27 companies) underwent factor and reliability analyses. Results showed that "soldier will" can be best represented by seven scales. Scales were internally consistent and highly intercorrelated. Scales bore significant and positive relationships with life and army satisfaction, and inverse relationships with duty stress, personal distress, medical problems, and wanting to get out of the Army.

Introduction

The U.S. Army has attempted to increase combat readiness through a series of initiatives known as the New Manning System (NMS). Traditionally, all soldiers are assigned to units as individuals. The NMS approach (called COHORT) assigns, trains, and deploys soldiers as intact groups during their first three-year enlistment in the Army. The common experience of basic and advanced individual training, transfer of personnel in groups, and low personnel turbulence within the unit during the initial first-term enlistment afford the opportunity to build strong interpersonal relationships. These interpersonal relationships are believed to have both individual and group effects pertinent to the soldier's will to fight, such as greater group identity, cohesiveness, esprit, mutual caring, sharing, and providing of emotional and instrumental support. The beneficial effects of these expected changes in interpersonal relationships have strong logical and intuitive appeal and are grounded in recent research findings (see Griffith, 1985; Leavy, 1983).

A crucial element in evaluating whether NMS has these intended human effects is defining the psychological phenomena that compose the soldier's psychological readiness to fight (collectively known as "soldier will"). The objective of this paper was to develop reliable and valid measures of "soldier will," in addition to demonstrate their utility by comparing soldiers from unit replacement units (COHORT) to soldiers from individual replacement units (nonCOHORT) on measures of "soldier will."

Method

Sampling of Units and Respondents Under Study

The sampling frame consisted of nineteen battalions of which ten were infantry, four armor, and five field artillery. In addition, 44 "independent" COHORT companies and their matched nonCOHORT companies were included in the sampling frame. COHORT and nonCOHORT units were matched on three criteria: type of combat arms unit (namely, infantry, armor, or field artillery), post location, and site of OCONUS rotation. The total 137 companies under study represented 20% of the Army's total unit strength. Data obtained from 27 companies (one-fifth of the sampling frame) were used for analyses in this paper. COHORT and nonCOHORT units in this subsample were comparable in number of companies, type of combat arms unit, and post location. Units included were five battalions, two COHORT battalions (one infantry and one field artillery), three nonCOHORT (one infantry, one armor, and one field artillery).

All soldiers in the five battalions under study formed the pool of potential respondents. Questionnaire administrators (civilian contractors) were asked to achieve at least an 80% response rate of personnel assigned to each company. The overall response rate was 77.3%.

"Soldier Will" Questionnaire Instrument

The "Soldier Will" Questionnaire (available on request) was a compilation of behavioral and attitudinal measures. Sections of the questionnaire pertinent to research objectives are described below.

Unit cohesion and morale. Three general scales were used to assess unit cohesion and morale. The first was a 19-item Unit Cohesion and Morale Scale developed by the Israeli Defense Forces (Gal, 1983). Scale items tap the soldier's perception of his(her) unit's cohesiveness and morale. Twenty-five items were included in in the second scale. Items asked soldiers to rate statements about pride in and importance of oneself, the unit, and the Army in general; unit "togetherness" or cohesion; unit morale; and unit leadership. These items were taken from the "Field Forces Questionnaire" developed by Army researchers during World War II to investigate attitudes of soldiers prior to and after the Normandy invasion (Stouffer, DeVinney, Star, & Williams, 1949). The third scale, the Company and Squad/Platoon Perceptions Scale, totalled sixty-five items. Items asked soldiers about their perceptions of the quality of relationships among soldiers; small unit interpersonal relationships; competency of officers, NCOs, and soldiers; and preparedness for combat. This scale has been used in several previous Departmental investigations and was shown to have high internal consistency and research utility. Respondents rated items on all three scales using a 5-point Likert-type scale, generally ranging from "strongly disagree" (scored as 1) to "strongly agree" (scored as 5).

Mental well-being. The 18-item General Well-being Scale was developed by Dupuy (1978) and was a measure of the respondent's general psychological well-being. Questions asked respondents about such things as being bothered by nervousness, losing control of thoughts, feelings, and behaviors, feelings of hopelessness, downheartedness, and loss of energy and vitality. Respondents rated scale items using a Likert-type scale.

Life and Army satisfaction. Thirty-eight items made up this section. Respondents were presented with "life areas," such as marriage, family life, health, and neighborhood, and too, "life areas" specific to the military, like the sponsorship program, company's leave and pass policies, and the unit's concern for families. Respondents rated their degree of satisfaction with each life area on a 5-point Likert-type scale.

Results

"Soldier Will" Measures: Scale Construction

The first phase of analyses was aimed at clarifying the constructs of "soldier will" by conducting a series of factor and reliability analyses of unit morale and cohesion scale items. Respondent ratings obtained on the Unit Cohesion and Morale Scale, Modified Field Forces Scale, Company Perceptions Scale, and the Squad/Platoon Scale were factor analyzed. First, each scale was factor-analyzed separately, and second, items of similar content from these scales were pooled and factor-analyzed.

Interpretability of emergent factors was better for factor analyses of the individual scales than for that of pooled items of similar content from the separate scales. Both conceptually and in terms of item content, items that loaded on some scales were very similar to those loading on others. To reduce redundancy and to develop a manageable number of scales measuring "soldier will," items that loaded on similar factors were combined to form one scale. When item content clearly overlapped, redundant items were eliminated. This procedure yielded seven measures of "soldier will:" (1) company combat confidence; (2) senior command confidence; (3) small-unit command confidence; (4) concerned leadership; (5) sense of pride; (6) unit social climate; and (7) unit teamwork. (Items comprising each scale and psychometric properties of each scale are available upon request.)

These new scales then underwent factor and reliability analyses in order to establish their unidimensionality and internal reliability. Overall, scales showed high internal consistency (respectively, for each of the seven scales, Cronbach's alphas = .91, .94, .87, .88, .84, .86, and .83). Scales that showed strong unidimensionality were senior command confidence, sense of pride, and unit teamwork. Company combat confidence comprised three factors: general combat confidence, confidence in weaponry, and confidence in oneself. Concerned Leadership consisted of two factors: concern of soldier welfare and personal contact with leaders. The Unit Social Climate Scale comprised three factors: trust and caring among soldiers, availability of instrumental support, and friendships among soldiers.

Validity of the "Soldier Will" Measures

If the "soldier will" scales tapped a broader, more unitary construct called unit esprit or group cohesion, then the scales should be highly interrelated. The first approach to validity (construct) intercorrelated the "soldier will" measures. Over one-half (12 of 21) of the intercorrelations among the "soldier will" scales had correlations of .60 or higher. Six other correlations were nearly .50 or higher. The remaining three correlations ranged from .37 to .40, and these were between the Senior Command Confidence Scale and Concerned Leadership, between Senior Command Confidence and Unit Social Climate, and between Senior Command Confidence and Unit Teamwork. Results supported the notion that "soldier will" scales tapped a broader, more unitary psychological construct of group cohesion.

The logic of the second approach to validity (concurrent) was: If "soldier will" scales assessed positive unit characteristics, then soldiers who report positive unit characteristics (as measured by "soldier will" scales) should also report greater positive life adjustment, less personal distress, fewer medical problems, and more willingness to reenlist. Results supported these expectations. Army satisfaction bore the highest relationship to the "soldier will" measures; four of the seven correlations were .57 or higher. The General Well-Being Scale bore the next highest relationship with the "soldier will" scales. The measures least correlated with "soldier will" was Life Satisfaction.

The "buffering effect" of unit esprit and cohesion on personal distress has a strong intuitive appeal and empirical precedent. Research investigating social support and their "buffering effect" of negative consequences of stress on personal well-being (see Leavy, 1983) strongly suggests that soldiers who report more cohesiveness among their unit members are less likely to experience the deleterious effects of stress and to report lower levels of personal distress and medical problems. Given this, inverse relationships are expected between measures of "soldier will" and personal distress.

All distress measures were significantly and negatively correlated with "soldier will" measures. The number of hours in a day and weekends in a month worked bore the highest relationships to small-unit command confidence, concerned leadership, and unit social climate. Of the time-spent-at-work variables, field time had lowest, and at times, no relationships with "soldier will." Of all the distress measures, not having enough time to attend to personal, family, and recreational needs was most strongly correlated with the "soldier will" measures. Seeing the doctor, taking medications for nerves, and worry interferring with work were all inversely related to "soldier will," especially to the soldier's sense of pride. Not being able to work because of worry was most strongly correlated with company combat confidence. Wanting to stay in one's unit after first-term enlistment and wanting to reenlist were positively correlated with the "soldier will" measures. Wanting to get out of the Army was significantly and negatively related to all "soldier will" measures.

Differences in "Soldier Will" between COHORT and nonCOHORT Soldiers

A series of multiple regression analyses were conducted in which "soldier will" measures served as the criterion variables and soldier and unit characteristics served as predictors. The purpose of these analyses was to ascertain the relationship of unit status--either COHORT or nonCOHORT--to each of the "soldier will" measures, while controlling for demographic and unit characteristics.

Although the amount of variance contributed by predictors to each of the "soldier will" measures was significant, the magnitude was relatively small (ranging from 3.6% to 11.5%). Noteworthy though is the greatest proportion of variance accounted for in four of the seven "soldier will" measures was whether the unit is COHORT (respectively for the Company Combat Confidence, Senior Command Confidence, Concerned Leadership Scales, and Unit Teamwork, .035/.075 ($F(1,2452) = 92.84$, $p < .01$), .019/.036 ($F(1,2449) = 48.35$, $p < .01$), .024/.038 ($F(1,1684) = 41.96$, $p < .01$), and .040/.074 ($F(1,2528) = 1366.49$, $p < .01$). For the other three "soldier will" scales, COHORT status was the second highest contributor to the variance in "soldier will" measures after rank or age. In addition, COHORT soldiers had higher scores than those of nonCOHORT soldiers on all "soldier will" scales when demographic and unit characteristics were held constant.

Units that differed substantially in training and performance were compared on the "soldier will" measures. Five paratroop COHORT companies from another data base were added to the present sample for this analysis. These five companies and the 27 companies in the current sample were placed in three categories based on their training, specialization, and perceived "eliteness:" Paratroop COHORT (most highly trained, specialized, and "elite"); COHORT (next most highly trained, specialized, and "elite"); and nonCOHORT (least trained, specialized, and "elite"). Means on each "soldier will" measure were calculated for each company within each category. To detect differences between company means across the three categories, a one-

way analysis of variance (ANOVA) was performed. Means of the company mean scale scores progressively increased from nonCOHORT to COHORT to paratroop COHORT consistently, though not always significantly. The three categories of companies significantly differed on four of the seven "soldier will" measures; these were Senior Command Confidence ($F(2,29) = 9.36$, $p < .001$), Small-Unit Command Confidence ($F(2,29) = 7.76$, $p < .01$), Concerned Leadership ($F(2,29) = 3.56$, $p < .05$), and Unit Social Climate ($F(2,29) = 7.18$, $p < .01$). Companies did not differ in their combat company confidence, sense of pride, and unit teamwork.

Discussion

Soldier attitudes were reliably measured, and based on item content, sets of attitudes were named as measuring specific aspects of "soldier will." "Soldier will" scales had high internal consistency and displayed both empirical and conceptual coherency as demonstrated by factor analytic results.

The validity of the "soldier will" measures was shown by their high degree of interrelationship. Although the scales measured different aspects of "soldier will," conceptually they are subsumed under a broader, more unitary psychological construct, perhaps called soldier morale, esprit, or will, and therefore, a high degree of interrelationship would be expected. Also, it was expected that "soldier will" measures should bear relationships to measures of life adjustment and personal distress. "Soldier will" was positively related to positive life adjustment (namely, life and Army satisfaction, and psychological well-being) and negatively related to personal distress, medical problems, and wanting to get out of the Army.

These scales were also expected to differentiate soldiers from units that have undergone special training and deployment designed to enhance "soldier will," namely COHORT. COHORT soldiers showed higher levels of "soldier will" than did nonCOHORT soldiers. In analyses comparing paratroop COHORT to COHORT to nonCOHORT companies on "soldier will," mean "soldier will" scores increased in magnitude from nonCOHORT to COHORT to paratroop COHORT companies consistently, though not always significantly. Despite the apparent success in achieving reliable and valid measures of "soldier will," the question that remains is: "How adequately do 'soldier will' measures translate into training performance, and more importantly, into combat readiness?"

References

Dupuy, H. J. (1978, October). Self-representations of general psychological well-being of American adults. Revision of a paper presented at the meeting of the American Public Health Association, Los Angeles, CA.

Gal, R. (1983). Unit Morale--The Secret Weapon of the Israeli Defense Forces. Paper presented at the Third International Conference on Problems of Stress. Tel Aviv, Israel.

Griffith, J. (1985). Emotional support providers: Who are they? Where are they met? and the relationship of network characteristics to psychological distress. Basic and Applied Social Psychology, 7, 41-60.

Leavy, R. L. (1983). Social support and psychological disorder. Journal of Community Psychology, 11, 3-21.

Stouffer, S.A., DeVinney, L. C., Star, S. A., & Williams, R. M. (1949). Studies in social psychology in World War II: Vol. 2. The American Soldier: Combat and its aftermath. Princeton, NJ: Princeton University Press.

Applying Qualitative Measures Over Time in Understanding Stress and Cohesion in the Combat Battalion

LTC T. Paul Furukawa, Ph.D., ACSW, Medical Service Corps

In depth longitudinal study of a single Army unit provided observations of both official ("frontstage") and behind-the-scenes ("backstage") behaviors that serve as indicators of unit stress and unit cohesion. The inclusion of a battalion being formed in the new Light Infantry Division offered two unique opportunities: (1) to study the same soldiers in a battalion comprising three COHORT rifle companies stabilized for three years, and (2) to compare naturalistic field-work findings (from interviews and participant-observation) with quantitative findings from periodic written surveys of issues relating to stress and cohesion in the same battalion.

Front-stage and back-stage information. Front stage information originates from the Public Affairs Office or from other forums that are "managed" for the purpose of conveying desired impressions, such as on-the-record interviews with leaders, official reports, public ceremonies, or meetings with representatives of higher headquarters. Impression management results in conveying perceptions about unit stress and cohesion (and other topics) that are indicative of the state of the unit but are incomplete. Back-stage information accounts for the preponderance of individual and group behavior, and includes off-the-record interviews, observations of informal gatherings, and participation in field exercises. Two techniques used to facilitate acquiring back-stage information were: (1) housing a research assistant in the barracks with single soldiers during site visits, and (2) participating in the life of the battalion (e.g., taking part in physical training, going to the field).

Qualitative and quantitative measures. Soldier surveys that are well-designed can provide useful measures of opinion and changes in opinion over time. Frequent contact with the battalion permited observation of events, discussion of the meaning of these events with various groupings of soldiers, and placement of these events within the context of the unit's life cycle. Comparisons between qualitative observations and quantitative survey results enriched the utility of both forms of data collection; likewise, survey results have pointed to areas requiring further observation.

Different Methods of Coping with Stress Among Hispanic Soldiers

Mary L. Lozano, Ph.D.
WRAIR, Department of Military Psychiatry

The general well-being of the Hispanic soldiers is an issue of importance to the U.S. Army since the Hispanic content of the force is expected to grow at a rapid rate. Hispanics are the second largest racial/ethnic group in the armed forces today. Due to their growth in numbers, there have been and there will continue to be an interest in the recruitment of Hispanics into the U.S. Armed Forces particularly since population changes are expected to occur that will affect the supply and demand for 18 to 21 year old volunteers in the near future. When compared to 1981, by 1987 there will be 1.2 million fewer men and women between 18 and 21 years of age, and approximately 4 million fewer by 1995.

The Army too has little information on health issues peculiar to Hispanic personnel. This is particularly true of those health problems and issues that may be stress related. Without sufficient information on Hispanic health issues, the Army Medical Department may not be able to respond adequately to many kinds of mental and physical disorders among Hispanic soldiers. Since the prevention of psychiatric casualties is dependent upon the efficacy of programs designed to maintain healthy military personnel, and since the rapid restoration of psychiatric casualties in combat depends in good measure upon understanding the symbolic realtionships of symptomatologies to stress, the needs of this particular group of soldiers become an important consideration for military health care providers.

Since little is known of the behavior concepts and values dealing with the health of Hispanics in the Army, the aim of this pilot study was to produce an ethnography of the Hispanic soliders' concepts of stress, illness and their method of coping with such stress within the structure of the military. Through the use of standard methods of of cultural anthropological analysis and techniques, information which addressed the stress issues of various Hispanic subethnic groups (e.g., Puerto Ricans, Mexican Americans, Cuban Americans, etc.) was collected in the form of case studies. A sample population was pulled from several Army posts from a random selection of combat units and combat support units. Along with a structured interview which contained health related questions, an unstructured interview was conducted which focused on the Hispanic soldiers' perceptions of military life. Results of thse interviews will be discussed.

Preliminary Report on Innovative Human Dynamics and Combat Effectiveness in the 7th Light Infantry Division

Faris R. Kirkland
Department of Military Psychiatry, Walter Reed Army Institute of Research

Abstract

A research program to assess the interactive effects of the COHORT system, family support programs, and leadership initiatives on combat effectiveness is in progress in the 7th Light Infantry Division. The program employs a longitudinal interactive combination of surveys and on-site observation. Preliminary research indicates that commanders in the 7th Light Infantry Division use supportive, empowering leadership. The light infantry mission emphasizes training in autonomous action. Synergistic interaction between these factors and the COHORT system has created human dynamics which lead to personal gratification and exceptionally high levels of combat proficiency.

Background

The research program described in this paper is designed to assess the interactive effects of three initiatives--the COHORT system, the Long Range Family Support Program, and the Chief of Staff's new leadership approaches--on interpersonal dynamics in Army units, and the effects of these dynamics on combat effectiveness.

Selvin in his pioneering work on interpersonal dynamics in military units differentiated between leader behavior and the followers' perceptions of that behavior. He defined the latter as "leadership climate" (Selvin, 1960:12-15). He also worked with the interaction between leadership climate and differential sensitivity to leadership arising from differing characteristics of subordinates (130-1). Selvin concluded that persuasive (non-coercive, confidence building) leadership was most effective (45). Post-Vietnam research associated the quality of human relations at the unit level with misconduct (Eckermann, Williams, & Ramsay, 1975), military proficiency (Novaco, et al., 1979) and retention (Kirkland & Good, 1979). The intensive investigation of attrition in the late seventies revealed the importance of commanders' attention to subordinates' personal and career needs (Goldman & Worstine, 1977), meaningful training and work (Zurcher, et al., 1979), efficient use of soldiers' time (Greenberg & McConeghy, 1977; Holz & Schreiber, 1977), honesty (Fullerton, 1984; Gottlieb, 1977; Porter & Steers, 1977), esteem for subordinates (Ingraham & Manning, 1980; Johns, 1983) and sense of belonging (Ingraham, 1984; Jacobowitz, 1980; Sodetz, 1983). Other investigators confirmed Selvin's conclusion that authoritarianism is not an optimal mode of military leadership (Mobley, et al., 1977; Moskos, 1977).

Scientists from the Department of Military Psychiatry at Walter Reed Army

Institute of Research, tasked to evaluate cohesion, social systems and soldiers' will to fight in COHORT units, discovered that the findings on leader behavior described above applied in new ways to COHORT units. COHORT units were relatively insensitive to a destructive (authoritarian, narcississtic, dependent--Dixon, 1976) leadership climate; the junior enlisted personnel performed well even when badly led. On the other hand, COHORT units were exceptionally responsive to positive leadership. Under concerned, candid, technically knowledgeable commanders COHORT units reached heights of competence and efficiency no conventional units could approach (Marlowe, 1985).

In 1984-5 the 7th Infantry Division at Fort Ord, California, underwent conversion to a light infantry configuration. Three infantry and one artillery battalion were 100 percent COHORT. The other six infantry and three artillery battalions had COHORT line companies or batteries and individual rotation headquarters and support units. The 7th Light Infantry Division offered the first opportunity to investigate the effects of the COHORT system fully developed and interacting with new family support programs and leadership initiatives.

Research Method

The research design is a longitudinal, interactive combination of surveys and naturalistic observation, with measures of combat-type performance as criterion variables. Surveys are administered five times over a three year period to all available personnel. The analyzed data from the surveys and reports of training performance orient the observation teams. Observers record language, content, and style of communications, the affect manifested, and the behavior resulting from the communications. The research team will combine data from surveys, observation, and performance measures to describe the leadership climate, social systems, communications, and competence in selected units. It is probable that causal relationships can be identified. These will be stated as hypotheses to be tested for general applicability in subsequent research.

The initial stage in the research was a reconnaissance performed in the spring and summer of 1985 to accustom the personnel in target units to the presence of observers, to test observational data collection modes, and incidentally to collect data. The results of this incidental data collection revealed that in leadership climate, social systems, and professional competence, units of the 7th Light Infantry Division differed significantly from other units in the U.S. Army.

Preliminary Findings

The leadership climate in the 7th Light Infantry Division and in most of its brigades and battalions is more supportive than in other comparable organizations. Most of the officers and NCOs in command and leadership positions treat their subordinates with respect. Superiors listen to their subordinates' ideas and in many cases collaborate with them in the development and realization of those ideas. Superiors express trust in their subordinates by providing them with complete and candid explanations, and by assigning missions for which their subordinates are allowed to work out the plans for execution. Superiors teach their subordinates to trust them by telling the truth, and by accepting bad news in a non-punitive manner. Attitudes toward mistakes, failures, and accidents are based on the realities of the situation, not on fantasies of zero

defects. Failure resulting from honest effort, development of subordinates, or experimentation is accepted as a legitimate price of progress.

A second aspect of the supportive command climate is concern for the professional development and personal and familial welfare of subordinates. In the 7th Infantry Division this concern is not paternalistic but collaborative. Development, promotion, and recognition of every soldier is a collective objective of every squad/section, platoon, company, and battalion. The impact on professionalism is spectacular. In one 9 month-old battery, almost every cannoneer is already qualified to serve as a section chief. In one 7 month-old infantry battalion, 145 men earned Expert Infantryman Badges.

With respect to welfare, a soldier's problem is the unit's problem--beginning at squad/section and progressing to division depending on which resources are required to resolve the problem. Families are integral--not subordinate--parts of companies and batteries. Family support groups, organized around emergent leadership, deal directly with the unit commander. A familial issue that distracts or disturbs a soldier is a matter of concern to his colleagues and to the chain of command.

The research team found the units of the 7th Infantry Division to be capable of reaching high levels of military proficiency in a very short time. One infantry battalion, and one artillery battery, went from activation through all squad and platoon level training and evaluations and completed company/battery training qualification tests (ARTEPs) in 90 days. All of the newly activated battalions pursued ambitious training programs at breakneck speed. The demand for challenging training springs from the junior enlisted personnel. The primary concern of commanders of those battalions composed of mature COHORT companies is providing progressively more sophisticated training throughout the lives of their units. One battalion, with COHORT companies that had been together for six to fifteen months, was airlifted to Korea in 1984 for maneuvers against Korean Army units. Within 48 hours after arrival in strange terrain, the battalion assumed a stay-behind mission. Its squads and platoons allowed themselves to be bypassed, then infiltrated into the rear of the Korean division. Cut off from resupply, the units evaded Korean patrols for three days then fell on the Korean logistical services and caused great confusion. Though none of the units of the 7th Infantry Division has seen combat, they all appear to have the ingredients that historically have made units cohesive and resilient in combat (Shils & Janowitz, 1948).

Discussion

In most of the battalions in the 7th Infantry Division the talents, energies, and interest of soldiers even in the lowest ranks are fully committed to the mission. Soldiers of all ranks respect each other as colleagues. Leaders at all levels feel secure taking the initiative, and zest for responsibility is real. Members of the 7th Infantry Division appear to feel a familial kind of affection for one another within and across rank strata. Concurrently, units in the division have achieved higher levels of combat proficiency earlier than other units, and are continuing to develop.

What appears to be taking place is synergistic interaction between the

COHORT system, the light infantry mission, and initiatives in leadership and family integration. The COHORT experience creates horizontal cohesion from which individual soldiers derive feelings of security, belonging, and identity; and develop shared experiences and values. Salient among the common values that unite the junior enlisted soldiers in all COHORT units is dedication to combat effectiveness. They respond to leaders whom they perceive as sharing their dedication, and as technically competent to enhance their military proficiency.

The light infantry mission and training emphasize the ability of small combat teams to survive and fight autonomously. Light infantry training gives junior personnel a sense of ownership and self-direction--which intensifies both their interest in developing their skills and their feelings of belonging and identity. They do not feel like pawns; they feel like significant actors in a nationally important endeavor. The appropriate type of command for troops destined to fight autonomously is the "power down" mode--decentralization of authority and discretion. This mode of command fosters junior soldiers' sense of self-worth and control.

Units commanded by the two commanders identified as the leading exponents of supportive, empowering leadership gave the most dramatic evidence of combat effectiveness. The company-level commander won a trophy last November for the fastest and most accurate gunnery among the fourteen comparable units in the division. The battalion-level commander is the one cited above twice--for taking his battalion through squad, platoon, and company ARTEPs in 90 days, and for qualifying 145 of his men as Expert Infantrymen.

Conclusion

A union of interested COHORT soldiers, a credible and dignifying mission, and empowering commanders constitute a positive circular process in which gratifications for all ranks are functions of commitment to the mission. The striking success of the battalions in the 7th Light Infantry Division owe something to the inherent small-unit autonomy of the light infantry mission, but autonomy is applicable to armor, artillery, and mechanized infantry units in any future air-land battle scenario. The crucial variable that has made most of the battalions in the 7th Light Infantry Division so extraordinarily effective is that most of the brigade-level and battalion commanders have had the psychological integrity to create a supportive, empowering command climate.

The central questions the research team will address in the next two years are:

One--Can the battalions of the 7th Infantry Division sustain the proficiency and commitment they have demonstrated to date?

Two--Do the company and battalion commanders in the 7th Infantry Division have professional or psychological characteristics different from those in other divisions?

Three--What are the precise dynamics by which the COHORT experience, the light infantry mission, and the leadership climate interact to produce extraordinarily effective units?

References

Dixon, N.F., On the psychology of military incompetence. London: Jonathan Cape, 1976.

Eckerman, W.C., Williams, J.R., & Ramsay, D.A. Observations of interrank conflicts at the company level: drug and alcohol abuse. RM 75-15. Alexandria: ARI, 1975.

Fullerton, T.D. "Comparing well-being and stress of several high-risk Army groups." Proceedings of the Ninth Psychology in the Department of Defense Symposium. Colorado Springs: U.S. Air Force Academy, 1984.

Goldman, L.A., & Worstine, D.A. Job satisfaction and reenlistment intent for first term personnel. Washington: USAMILPERCEN, May 1977.

Gottlieb, D. Attrition: The absorption and integration of newcomers. H. Wallace Sinaiko, ed. First term attrition, Volume I: Papers. Washington, DC: Smithsonian Institution, Jun. 1977. Referred to hereafter as (Sinaiko.)

Greenberg, M.G., & McConeghy, G. Exploratory development research of U.S. Navy-Marine Corps personnel - Phase I: factors affecting attrition. (Sinaiko).

Holz, R.F. & Schreiber, E.M. Increasing the retention of Army volunteers: meaningful work may be an answer. (Sinaiko).

Ingraham, L.H. The boys in the barracks. Philadelphia: Institute for the Study of Human Issues, 1984.

Ingraham, L.H., & Manning, F.J. Cohesion in the U.S. Army: who needs it, what is it anyway, and how do we get it to them? Paper presented at the Inter-University Seminar on Armed Forces & Society, Chicago, Oct. 1980.

Jacobowitz, D.W. "Alienation, anomie, and combat effectiveness. Air University Review, 31, 6 (Sept.-Oct. 1980) 22-34.

Johns, J.H. The changing nature of the American military and its impact on cohesion. Paper presented at the Northeast Regional Conference of the Inter-University Seminar on Armed Forces and Society, Apr. 1983.

Kirkland, F.R., & Good, J. Army reserve component retention. Philadelphia, PA: University City Science Center, Jul. 1979.

Marlowe, D.H. New manning system field evaluation. TR-1. Washington: Walter Reed Army Institute of Research, Nov. 1985.

Mobley, W.H., Hand, H.H., Logan, J.E., & Baker, R. Pre-recruit training values, expectations, and intentions of Marine Corps recruits (TR-2). Columbia, SC: Center for Management and Organizational Research, University of South Caroline, May 1977.

Moskos, Jr., C.C. It's a new ball game: changing expectations of military service. (Sinaiko).

Novaco, R.W., Sarason, I.G., Cook, T.M., Robinson, G.L., & Cunningham, F.J. Psychological and organizational factors related to attrition and performance in Marine Corps recruit training. Arlington: ONR, Nov. 1979.

Porter, L.W., & Steers, R.M. Organizational commitment and personnel attrition. (Sinaiko).

Selvin, H.C. The effects of leadership. Glencoe, IL: Free Press, 1960.

Shils, E.A., & Janowitz, M., "Cohesion and disintegration in the Wehrmacht in World War II," Public Opinion Quarterly, 12 (Summer 1948) 280-315.

Sodetz, F.J. A research perspective on drug and alcohol use in the Army. WRAIR-NP-83-3. Washington: Walter Reed Army Institute of Research, 1983.

Zurcher, L.A. Patton, J.F., & Jacobsen, A.M.B. The United States Naval Reservist: some observations on satisfaction and retention. Armed Forces and Society, 1979, 6, 82-110.

TOPIC: The Impact of Employment on the Psychological Well-Being of Army Wives: A Longitudinal Survey Study

AUTHOR: Jeannette R. Ickovics and Maj (P) James A. Martin, Ph.D.

Abstract

Information from 278 spouses of Army service members participating in a three-year panel study of the impact of the Army's New Manning System on family members is presented. The typical women surveyed was a young, recently married or remarried spouse of an enlisted soldier serving in a combat arms unit. Measures take 6 months apart suggest that the personal well-being of respondents was affected by their employments status. Specifically, these data show that those women who were not employed at time one, but subsequently attained employment, demonstrated the largest overall increase in general psychological well-being (GWB). On the other hand, those women who were employed throughout both points in time, had the sharpest decline in GWB. These results suggest that obtaining employment may enhance general well-being for military wives. At the same time for the military wife who must cope with the demands of home and family while receiving very limited assistance from her soldier-husband, the additional demands of employment may eventually have a negative impact on there well-being. Institutional ways of lessening the impact of multiple role demands on military wives are needed. It is also necessary to examine other contributing variables (e.g. social support), and to continue to monitor further changes in the general well-being of this sample as a function of their employment status over time.

The Military Family Research Program: An overview

Kathleen Saczynski, Ph.D.

A five year program in military family research has been initiated by the Department of Military Psychiatry, Walter Reed Army Institute of Research. The program is designed to investigate the relation between the military organization and quality of life for the military family. Understanding the nature of this relation, including the impact of the military structure on the family and reactions of military families to their environment is also critical to understanding certain problems or stresses in the military member who is the interface between the military and family systems.

In order to accomplish this program, the initial phase focuses on in-depth, exploratory, ethnological and systems oriented studies designed to identify the needs of military families, those stressors and supports within the military organization affecting family members, and particular coping strategies that may be critical for the stability and quality of life of the family. An important part of this phase is identifying the availability, use, and efficacy of support systems for family members. The presentation will discuss the initial phase of this research program.

Family Violence in the Military

Christine Russell
Walter Reed Army Institute of Research

Abstract

Increased attention is being given to the subject of family violence in the military. A study of spouse and child abuse at a large Army post was conducted. The study, based on 260 cases of child abuse and 245 cases of spouse abuse, indicates that stress factors present at the time of the incident were family discord, financial problems, sexual problems, and alcohol and drug dependency. Substance abuse and service connnected separation emerge as two consistent stressors in both the child and spouse abuse data which were extracted from case records. The preliminary study indicated that there is a need to investigate processes underlying child and spouse abuse in the military and to address the question of whether or not factors that contribute to family violence are specific to the military or outgrowths of other phenomena in the larger society.

Introduction

Background

A variety of factors contribute to family violence. The variables may be divided into those existing within the family and those acting on the family from outside (Straus, 1979; Gelles, 1979). Early research investigated the personality of the parent and established certain common characteristics such as low self-esteem, poor physical health, low frustration tolerance leading to aggression and previous abuse as children (Gelles, 1979).

Environmental factors related to abuse are listed as the family's specific life situation in terms of financial status, employment, social integration, family relationships, and general stress level (Schnall, 1977; McCullah, 1978). General community welfare including both cultural values and assumptions also influence the incidence of family violence (Finkelhor, David et.al., 1982).

Factors repeatedly referred to in research on the military concerning spouse and child abuse are: job related pressures and rank structures, mobility, deployment and other service connected separations, financial problems, low income, isolation from extended families and support structures, substance abuse, and bicultural marriages. (LaGrone, D.M., 1978; Helfer, R.E., 1974; Neidig, P.H., 1983)

A major factor necessitating a close analysis of family violence in the Army is that a growing percentage of the military, including new recruits, are married and have young children. In today's Army, with one million full-time and part-time forces, over 50 percent of the troops are married and younger than ever before. (Soldier Report II, 1983)

Methodology

The study included all families who were identified to Social Work Service on a large Army post as abusive or neglectful. The study is based on information collected between September 1983 and June 1984. The data were collected by social work staff members after training sessions were conducted by the researcher. The data itself was extracted from the Family Advocacy Case Management Team Incident Reports.

Results and Discussion

The study yielded a set of demographic statistics which provided a profile of child and spouse abuse. The data on child abuse indicated that military sponsors were mainly from the rank of E2 - E7, however, the majority of cases occurred in the E4 and E5 ranks. Thirty-one percent of the families lived on post while 67 percent lived off-post. The children were evenly distributed by sex. Sixty-two percent of the children were white; 31 percent were black and 7 percent were from other ethnic groups.

Additional data described type and nature of injury, relationship of abuser to child, who reported the abuse case and what services were utilized. It was found that 68 percent of the perpetrators were fathers; 29 percent were mothers; 3 percent were others, e.g., a grandparent or uncle. The majority of cases were reported by neighbors. Only 5 percent of the cases came through the emergency room. In terms of services utilized, medical services, mental health and counseling services were most often utilized on post.

Implications of this data must be drawn with caution. For example, E4 and E5 ranks comprise a large proportion of the population. There is also possible reporting bias which occurs when lower ranking individuals are channeled through social work services while higher ranking persons can consult a civilian agency. Since the data came from the post social work service, it is not possible to know how much reporting bias occurred. Taking these factors into consideration, it can still be noted that middle enlisted ranks would be a possible target population for preventive strategies.

There are trends which indicate that higher abuse rates occur in certain units than other units. Since 68 percent of the perpetrators were fathers, further investigation of the father's relationship to the unit should take place to determine if long duty hours, field training exercises, recurrent relocations, separation from family and friends, and threat to life have a relationship to abuse. If this is the case, certain preventive or interventive strategies could be implemented. For example, it can be found in the data that there were more abuse cases in a unit which had recently returned from a 6 month deployment. If some classes were held both

before and after deployment which alerted families to the possibility of disruption due to deployment, some abuse might be prevented. Families could also be advised of services available to them.

Neglect cases sometimes occur during deployments because young mothers do not receive their husband's paycheck. The mother is left with little or no resources for herself and her children. Better knowledge of such services as Army Community Services might circumvent some of these neglect cases. Since the data also indicated that the most often used services were social work or mental health services, it is apparent that Army Community Services could be better utilized by certain clients, particularly neglect cases. This would assist in removing some of the backlog that the counseling services on post are forced to deal with at present. Since only 5 percent of the cases were referred by the emergency room, it is possible that an additional interventive measure is necessary, e.g., further education of medical personnel and technicians about symptoms of abuse.

In the case of spouse abuse, the study indicated the 92 percent of the perpetrators were husbands; 2 percent of the abusers were wives and 6 percent were "others" (boyfriends). Eighteen percent of the cases are white; 62 percent are black and 20 percent are of other ethnic groups. Fifty-six percent of the cases occurred on-post; 44 percent happened off-post. Seventy-four percent of the perpetrators were from the E4 to E6 pay grades. The majority of cases involve first marriages (75 percent). Only 2 percent of the spouse abuse cases were serious enough to require hospitalization. Thirty-nine percent of the cases were, however, serious enough to require some treatment. Over one-half of the incidents were reported to social work service by military police. A higher percentage of spouse abusers than child abusers said that they had used drugs or alcohol prior to the event.

Again, as was the case with child abuse data, implications must be carefully drawn. While there may be some reporting bias and a large population of E4 to E6 personnel on the post, 74% is still a notable figure.

Rank is a factor which permeates the military environment. It carries financial benefits and also status, respect, authority and responsibilities. It is plausible that there is some correspondence of low rank to such social-structural elements as financial stress, crowded living arrangements and less control over one's life; all factors which have been suggested as contributors to family violence (Newberger, 1983).

As was the case with the child abuse data, there is some indication that there were higher rates of spouse abuse in some units than others. This suggests further research which would explore the effects of family separation due to military deployment on the occurence of spouse abuse. Consistent stressors in both the child and spouse abuse data were substance abuse and service connected separation.

Implications for Future Research

This study presents a demographic profile of abuse and suggests implications for program change; it also stimulates questions which only additional research could explore. The following issues should be addressed in future research:

* the special role of military stresses as possible contributors to family violence; reasons why some families are drawn closer together and others deteriorate

* whether stress in abusive families is greater than stress in nonabusive families or whether it is the coping strategies utilized by the families that differ

* whether there are fewer incidents of family violence in families who strongly identify with the Army and the unit.

* the relationship between substance abuse and family violence

Gelles, Richards J. and Straus, Murray A. (1979) Determinants of Violence in the Family: Toward a Theoretical Integration. Contemporary Theories About The Family. New York: Free Press.

Helfer, R.E. (1974) A Report of a Symposium on Child Abuse and Neglect in the Military. Amercian Medical Association CP, Annual AMA Mtg on Child Abuse/Neglect in the Military.

La Grone, D.M. (1978) The Military Family Syndrome. American Journal of Psychiatry. Vol 135: 9, 1040 1043.

Mc Cullah, F.D. (1978) Effects of Family Dysfunction of Military Operations: Mental Health Needs. USN Bureau of Medicine and Surgery, Washington DC BK, The Military Family and the Military Operation; DTIC ADA080943.

Newberger, Eli H., Newberger, Carolyn Moore, and Hampton, Robert L. (1983) Child Abuse: The Current Theory Base and Future Research Needs. American Academy of Child Psychiatry.

Neidig, P.H., Friedman, D.H. and Howell, W.L. (1983) Spouse Abuse Intervention Program in the Military Setting. USMC Recruit Depot, Parris Island, SC. UM, Behavioral Science Assn. Beaufort, SC.

Schnall, S.M. (1977) Characteristics and Management of Child Abuse and Neglect Among Military Families. USA, Project Care, BAMC, Ft Sam Houston, TX 78234. Children of Military Families. Hunter, Nice (EDS), GPO #008040001814.

NEW DIRECTIONS FOR FAMILY SUPPORT GROUPS IN A RAPID DEPLOYMENT COHORT FORCE

Joel M. Teitelbaum, Ph.D., Department of Military Psychiatry

Family Support Groups (FSGs) constitute a fast-developing social support phenomenon at Army installations. FSGs are quasi-voluntary family self-help organizations initiated by unit commanders, but organized and operated by soldiers' wives as mutual assistance networks for unit family members. Their assigned function at the company and battalion organizational levels is to help family members moderate the adverse effects of military stress during TDY absences and in case of unit deployment or casualties. In our study of COHORT infantry units of a new Rapid Deployment Force at a major U.S. installation the efforts of FSG volunteers were observed to exceed the expectations of the military chain of command.

FSG volunteers express an ethos of common identity as Army family members. Their activities minimize soldier rank distinctions, belying concerns that a "wives' chain of command" will result in which spouses "wear the husbands' rank." Volunteers extended caring social support to one another and provided outreach information to newly arrived married soldiers and family members about installation human resource agencies and unit military schedules. As the division entered into readiness status, company and battalion FSGs organized pre-deployment briefings for family members to assist in raising personal coping capacity during the soldier-husband's anticipated absence. The COHORT unit Family Support Groups took on additional roles that provided continuity during periods of time when soldiers and units were in garrison. They tried to incorporate soldiers living in the barracks into their growing exchange networks. They reached out to non-resident wives and family members of single soldiers through newsletters and written information. Family Support Group efforts heightened the awareness of commanders and COHORT unit soldiers about family needs. Military recognition of the value of stable families for COHORT unit cohesion and combat readiness has been stimulated by their activity. In today's Army there is a high ratio of married soldiers and much institutional concern about spousal and and child issues. Yet professional assistance resources are limited and not very responsive to the felt needs of family members and soldiers. Under these conditions a self-help,'lay' association that has a community identity and can respond caringly to human distress, such as the Family Support Group, offers positive new directions for stress management and better quality of life to Army families.

Toward a Component Model of Instructional System Design

Joan K. Gallini and Arthur D. Fisk
University of South Carolina

Abstract

Development of effective instructional programs is usually approached independently by two disciplines: cognitive psychology and instructional design. The lack of communication between the two fields has left unanswered many fundamental questions regarding training of skills. The present paper presents an initial, theoretical effort to bridge the gap between these two fields. A novel approach to instructional design is proposed. The training approach combines instructional design technologies and psychologically-based skill-training technologies derived from automatic/controlled processing theory.

It would be hard to find an institution in our society - be it education, industry, government - that is not concerned or affected by training. To clearly appreciate the global concern, readers need only consider a sampling of instances, such as the immediate training needs of the FAA in 1982 when the air traffic controllers were fired. Or the urgency felt today by banks, retailers, and other businesses to rapidly train employees to use new technology. Also, think about the well publicized concern that our high schools are graduating students who lack knowledge of basic skills.

Though on the surface the above situations appear to represent different problems, there is a theme common to all of them: The need for efficient formalized training to develop intellectual skills. Resolving the problem requires learning how to quickly move learners to a prespecified skill level. While the issue may appear commonplace, the solution to achieving the goal has eluded researchers and practitioners.

Generally, training programs are unguided in their development stages by theories of how humans learn and by models describing how instruction for such programs should be systematically formulated. Rather, they are directed by global and ambiguous objectives such as "students will be able to identify parts of speech" or "employees will be accurate when inspecting the product" with little or no theoretical basis for determining the components of the instruction. As a consequence, much time, money, and effort is expended in efforts to teach/train; moreover, to retrain and reteach individuals. More alarming than the cost of training is the general overall ineffectiveness of it. As Vreuls and Obermayer (1985) have indicated, trainers usually feel successful if they can train people to a minimum level of acceptable performance. They leave mastery to "on-the-job" training. There is an inherent danger in this training philosophy: Do you want to be under the care of a "minimally trained" physician? Indeed, society pays a high price for unnecessarily long training-retraining programs, high drop-out rates, and poor performance of trainees.

Development of effective instructional programs has been, traditionally, approached independently by two disciplines: 1) cognitive psychologists, who

have tried to apply knowledge about how people process information; and 2) instructional design technologists who have tried to determine the overall structure of the instruction for effective skills acquisition (e.g., how the instruction should be broken down into teachable units, how these units should be sequenced, etc.). However, the lack of communication between the two fields and the instructional technologists' use of an incomplete model of human cognition has left unanswered many fundamental questions regarding the training of skills.

This paper represents an effort to bridge the gap between the two fields by proposing an instructional design model that should predict when learning of task components will occur and by suggesting which components of a task should receive training. The proposed "decomposition" model provides a framework to guide task analysis when designing instruction of intellectual skills and is based on the interactions between cognitive mechanisms mediating both the learning and the performance of a given task and the nature of the task (see below).

Theoretical Framework. Human information processing can be categorized as two qualitatively different processes called automatic and controlled. Current attention research in psychology assumes that behavior results from a continual interaction between automatic and controlled processes (for example, see Fisk & Schneider, 1983; Logan, 1985; Schneider & Shiffrin, 1977, Shiffrin & Schneider, 1977). Automatic processing develops after a behavior has consistently occurred many times. Automatic processes are fast, occur effortlessly, not limited by attentional capacity, not under direct conscious control, and very difficult to modify once developed. Manipulating a fork at dinner is an example of automatic processing. Controlled processing requires little experience to develop, is easy to modify, is slow, effortful, and places severe limits on attentional capacity. Remembering a phone number long enough to dial it is an example of controlled processing. The functions of controlled processing are to: develop new automatic processes (i.e., allow learning to occur), deal with novel or inconsistent situations, maintain information in temporary memory, activate automatic processes, and inhibit or modify existing automatic processes.

The distinction between automatic and controlled processing has demonstrated that statements such as "practice makes perfect" are only partially correct. Dramatic changes in performance (i.e., a change from controlled to automatic processing) occur only when individuals are given extensive consistent practice. Consistent practice occurs when stimuli and responses are consistently mapped (CM); that is, across training trials the individual makes invariant overt (or covert) responses to stimuli (or classes of stimuli). If individuals receive varied mapping (VM) training - that is, a given stimulus requires responses that unpredictably change across time - automatic processing will not develop and performance will not dramatically change with practice.

The automatic/controlled perspective provides an interpretation of when and why training will be effective. Research has demonstrated that a major role of practice is to allow trainees to focus on and learn task consistencies so that automatic component skills develop (e.g., see Schneider & Fisk, 1983). In order to perform complex intellectual tasks such as reading, programming, or problem solving many subskills must be automatized so that attention can be

devoted to higher-order cognitively demanding aspects of the total task (e.g., Anderson, 1983; Fisk & Schneider, 1983; Resnick & Ford, 1981). It is not sufficient that consistent subcomponents be learned; rather, they must be trained to a level of automaticity so that performance on the cognitively demanding task components is optimal. (For example, imagine trying to read if you had to continuously monitor the shape of each letter of each word.)

Impact on instructional design systems. Such theory-based findings about cognitive processes provide significant input to the efforts of instructional design scientists to identify how instruction for training should be developed into teachable components. Automatic/controlled processing theory provides a strong theoretical and empirical foundation for the hypothesis that skill acquisition will be facilitated if instruction is broken into a sequence of consistent trainable components. These consistent components will foster automaticity thus "freeing up" mental resources for integrating information, building more complex processes and other activities requiring controlled processing resources.

Current instructional design systems have neglected the considerations of automatic/controlled processing theory. Typical systems have relied solely on a hierarchical notion that "lower-level" knowledge must be acquired before "higher-level" knowledge can be acquired. However, these system have not provided an operationally effective means for identifying why lower-level knowledge is lower-level, how lower level-knowledge is built up into higher-level knowledge,etc.

More recent theoretical approaches to instructional design have attempted to describe "procedural" aspects of cognitive tasks. For example, Landa's (1983) algo-heuristic theory of instructional design argues for breaking complex cognitive tasks into elementary cognitive processes. Unfortunately, this approach, like other instructional design systems, is theoretically based with no empirical evidence of how to identify elementary processes or how to break complex instruction into elementary components. Furthermore, and as serious of a concern, in current approaches to instructional design an appreciation of both how and why human information processing changes as skills develop is consistently missing.

The present approach to instructional design systems attempts to overcome the limitations described above. The model is used to decompose complex instruction into trainable consistent components. The decomposition model incorporates aspects of automatic/controlled processing theory and relevant aspects of previous instructional design systems (e.g., Merrill, 1983). The decomposition process creates a 2 X 3 X 3 matrix of "cells"; each cell represents a potential category of task components. Each cell in the model defines and predicts the amount of learning (automatization), the type of learning, and the "usability" of that learning for higher-level components/skills. Each cell is a combination of: 1) the degree of component consistency; 2) the intended learning outcome (e.g., remember facts, use rules, or extend rules in problem solving); and 3) the nature of the task (e.g., fact, concept, principle, procedure). Concern is also given to the amount of cognitive resources available/demanded of each component. Once the task is decomposed into its consistent trainable components then training guidelines, derived from automatic/controlled processing theory (Schneider, 1985), can be applied to the training situation.

Abbreviated example of the decomposition process. Although there is insufficient space in the present paper to completely describe the workings of the model, a simple illustration of the decomposition process will be presented. Imagine an instructional situation requiring students to learn to program in BASIC. More specifically, imagine developing a lesson focusing on the concept of "variables". Application of our model would provide a systematic approach to reducing the complexity of the instruction by decomposing it into as many consistent elements as possible. The first step is to describe the instructional content in terms of: the nature of the task (concept learning); the expected learning outcome (students should apply rules); and the degree of consistency. At the single instance level, learning the concept of "variables" involves, what appear to be to the novice, many inconsistencies; thus, the lesson must be further decomposed.

At this point, and in search of consistency, the decomposition process should focus in on the nature of the task (such as fact, concept, etc.). With this focus, it is apparent that to teach the concept "variable" for both understanding and usage the following must be considered: A definition must be learned (fact knowledge, consistent). Critical attributes should be taught (concept, apparent inconsistencies). Note, this is an apparent inconsistent component since attributes differ from one application to the next. Therefore, our model would suggest training a small subset of high frequency exemplars of the category "variable" (concept level, consistency at the "category" level). Finally, the learners must be able to apply an operational rule of "variable" (procedure, consistency at the rule level). To do this, consistency must be maintained at the rule level and the student must be given many opportunities for correctly applying the operational rule.

Further tuning would be required even for this simplistic example of teaching the concept of variables. However, the reader should see that the lesson would be composed of three primary consistent components. The practice sessions would consist of many correct executions (trials) emphasizing each of these components in turn.

The potential significance of the current decomposition model is two-fold. First, a research effort aimed at exploring the limits of the above decomposition process has the promise of producing objective, general principles for decomposing complex intellectual skills. Being able to specify consistent components of intellectual skills is paramount to developing superior training programs. The methods that are used today to develop training of complex intellectual skills are often derived from instructors' or experts' personal experiences. Being personal experiences, these "training technologies" are often intuitively understood. Intuitions, by their very nature, are hard to consistently pass on to other, less informed or experienced, instructors. Second, by focusing research on instructional design systems within the framework of an empirically solid cognitive theory of skill acquisition, we have the promise of general principles to guide us in the further refining of training technologies. Often, when using the instructional design systems that are in place today, if we refine an instructional training program (by guess work, experimentation, etc.) we can usually only guess what the next step is for further refinement. The present system helps eliminate that unfortunate problem.

References

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Fisk, A. D. & Schneider, W. (1983). Category and word search: Generalizing search principles to complex processing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 177-195.

Landa L. N. (1983). The algo-heuristic theory of instruction. In C. M. Reigeluth (ed.) Instructional-design theories and models: An overview of their current status. Hillsdale, NJ: Erlbaum.

Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. Canadian Journal of Psychology, 39, 367-386.

Merrill, M. D. (1983). Component display theory. In C. M. Reigeluth (ed.) Instructional-design theories and models: An overview of their current status. Hillsdale, NJ: Erlbaum.

Resnick, L. B. & Ford, W. W. (1981). The psychology of mathematics for instruction. Hillsdale, NJ: Erlbaum.

Schneider, W. (1985). Training high-performance skills: Fallacies and guidelines. Human Factors, 27, 285-300.

Schneider, W. & Fisk, A. D. (1983). Attentional theory and mechanisms for skilled performance. In R. Magill (Ed.) Memory and control in motor behavior. New York: North-Holland Publishing Co.

Schneider, W. & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review, 84, 1-66.

Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.

Vreuls, D. & Obermayer, R. W. (1985). Human-system performance measurement in training simulators. Human Factors, 27, 241-250.

Armor Crewman Performance on a Tank Gunnery Video Game Under Varying Conditions

Millicent H. Abel

U.S. Army Research Institute-Fort Knox

Abstract

The purpose of this research was to examine the nature of the learning curves on a tank gunnery video game in a series of three experiments. Accuracy measures were collected over trials on experienced and inexperienced groups in Experiments 1 and 2. Subjects used the Gunner's primary sight in Experiment 1 and the Gunner's secondary sight in Experiment 2. Significant improvement was indicated in accuracy over practice trials with no differences found in performance between groups in the first two experiments. Subjects' accuracy and speed were examined over trials under two different game formats and two different target kill zones in Experiment 3. Results of Experiment 3 indicated a significant difference in accuracy between the standard kill zone and reduced kill zone groups with the reduced kill zone groups being less accurate overall. Improvement in accuracy was demonstrated using a revised video game format with the distribution of lives and ammunition into separate games. No improvement was indicated using the standard game format. Improvement was found for all groups in the average time to fire.

INTRODUCTION

The development of inexpensive yet effective means of training and sustaining proficiency levels of crewmen's tank gunnery skills have high priority in Army research. Due to the rising cost of equipment and ammunition, training devices and simulators are being utilized as a means of supplementing live fire exercises. One such device, the Battlesight tank gunnery game, was developed for use by Armor crewmen in barracks dayrooms and leisure areas. The Battlesight models a single-player video game found in many commercial arcades. The appealing nature of a video game format was utilized to voluntarily increase crewmen's gunnery practice. Limited programmability allows the suppressing or altering of several Battlesight parameters. For example, game play can be programmed to use either the M60A1 tank gunner's primary sight (M32E periscope) used in the normal mode or the secondary sight (M105D telescope) used in the degraded operational mode. Characteristics of the Battlesight game, such as the realistic reticle (sight) simulation and the mirroring of crewmen's tank gunnery skills required for game play, have generated considerable interest in the potential training value of this device.

Minimal research has been conducted concerning the use of video games in skill acquisition. Skill acquisition involves two main components, speed and accuracy. The training value of any device depends upon the demonstration of improvement in these components. Therefore, the first two experiments examine the learning curves defined by accuracy over trials. Experiment 3 examines the question of speed versus accuracy. In standard video games, similar to the Battlesight, the avoidance of being destroyed by shooting first requires the subject to demonstrate speed. Many video games may subtly emphasize speed of response at the expense of accuracy. Massed versus distributed practice is examined in Experiment 3 as it relates to speed and accuracy. Massed practice is the practicing of some skill over repeated trials without rest. Distributed practice is the spacing of practice trials with periods of rest. Research on motor skills has reported attenuated performance effects when the massed practice approach was used while distributed practice led to heightened performance. The massed practice model could be analogous to the video game format of supplying several lives and an abundant ammunition supply within one game. For example, one life could be considered one trial. If several lives are allowed for game play, then several repeated trials without rest would occur in the practice sequence. Therefore, the massing of lives and ammunition in video games may increase speed at the expense of accuracy while the distribution of lives and ammunition into separate games may increase accuracy along with speed.

Experiments 1 & 2

Method

In Experiment 1, twelve tank crewmen, fired 500 rounds of ammunition using the Gunner's primary sight (M32E periscope). These crewmen were divided into an experienced group (Tank Commanders/Gunners) and an inexperienced group (Drivers/Loaders). The 500 rounds of ammunition were divided into 50 round trials to examine practice effects. Accuracy was measured by the total number of hits and first round hits per trial. In Experiment 2, twelve tank crewmen, fired 200 rounds of ammunition using the Gunner's secondary sight (M1050 telescope). These crewmen were also divided into groups based on duty position. The 200 rounds of ammunition were divided into 50 round trials with accuracy being measured as in Experiment 1.

Results

No significant group differences were found in accuracy in either experiment. However, significant practice (trial) effects were found for number of hits ($p<.01$) and number of first round hits ($p<.01$) in both experiments with significant positive linear trends ($p<.01$) in the trial means. Figures 1 and 2 illustrate the linear trends for number of hits in Experiments 1 and 2. The trends for number of first round hits were similar. These figures display the significant improvement in accuracy for both groups when using the Gunner's primary and secondary sights on the Battlesight.

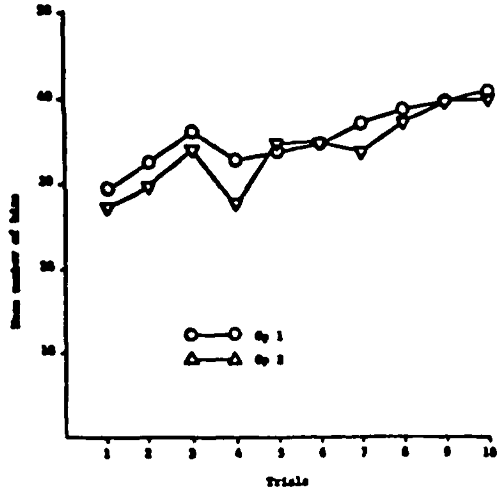


Figure 1. Mean number of hits per trial for Experiment 1

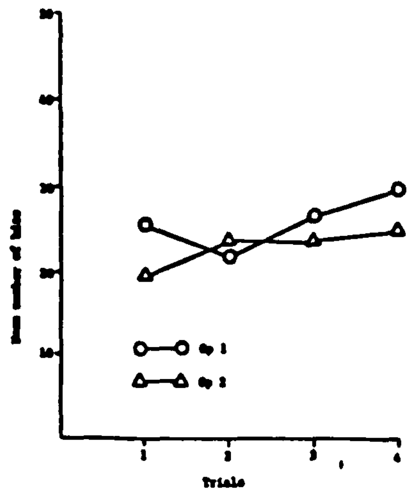


Figure 2. Mean number of hits per trial for Experiment 2

Experiment 3

Method

Sixty soldiers nearing graduation from one station unit training and trained on M60A1 tank gunnery skills were assigned to one of four experimental groups. The experimental groups received practice on the Battlesight under four game configurations: (a) Group 1: Battlesight with the number of lives and ammunition distributed and standard kill zone, (b) Group 2: Battlesight with the standard video game configuration of massing lives and ammunition and the standard kill zone, (c) Group 3: Battlesight with number of lives and ammunition distributed and a reduced kill zone, and (d) Group 4: Battlesight with the standard video game configuration and a reduced kill zone. The distribution of lives and ammunition in two of the experimental groups was accomplished by programming each trial into three game blocks. Each game block consisted of one life and 20 rounds of ammunition. The video game version with the massing of three lives and 60 rounds of ammunition existed in the other two experimental groups. Each game block or trial ended when the subject's tank(s) were hit or all ammunition was expended. The kill zone can be represented by a rectangle totally surrounding a threat tank. The kill zone was reduced from 100% in Groups 1 and 2 to 50% in Groups 3 and 4. At 100%, the threat tank could be destroying by hitting anywhere within that rectangle. At 50%, the rectangle is decreased by one-half which significantly reduces the area in which a hit could destroy a threat target. A data collector recorded the measures at the end of each game block or trial. The measures recorded were: (a) elapsed game time (b) total hits and (c) total rounds fired. In Groups 1 and 3, measures were recorded each time the subject was "killed" one time or expended all ammunition. This signaled the end of game block 1. The game block was repeated two more times for the total trial performance. In Groups 2 and 4, the subject continued playing until he was "killed" three times or expended all ammunition which constituted the end of the trial. Measures were subsequently recorded. The subject completed three total trials in one session.

Results

Analysis of variance on repeated measures was performed to de ·i·. if significant differences existed for game configuration and kill zone . . centage of hits, number of rounds of ammunition fired and the average ..... to fire. Because all subjects fired varying numbers of rounds of ammunition, percent hits were used as measures of accuracy. The average time to fire variable was computed by dividing the elapsed game time by the number of rounds of ammunition fired which resulted in the average amount of time to fire one round of ammunition. This variable was used as a measure of speed.

There was an overall significant effect for kill zone on percent hits $p<.001$). The kill zone by trial interaction was also significant ($p=.045$), suggesting a difference in the learning curves for the two kill zone groups across trials, see Figure 3. There was no significant overall game configuration effect nor was there a significant game by trial interaction on percent hits. However, the linear component of the game by trial interaction approached significance ($p=.065$). This result indicated a possible difference in the linear components of the trends of the trial means for the two game configuration groups, see Figure 3. There was no overall game configuration by kill zone interaction nor was there an overall trial effect on percent hits. When testing for simple effects of trial within kill zone, a significant positive linear trend in percent hits was found for the standard kill zone groups ($p<.05$). When testing for simple effects of trial within game configuration, a significant positive linear trend in percent hits was found for the revised video game groups ($p<.05$). There were no overall significant group differences on the number of rounds fired. The overall trial effect was significant ($p<.001$) with a significant positive linear trend of trial means ($p<.001$), see Figure 4. There were also no significant group differences on the average time to fire. However, a significant trial effect was found ($p<.001$) with a significant negative linear trend in trial means ($p<.001$), see Figure 5.

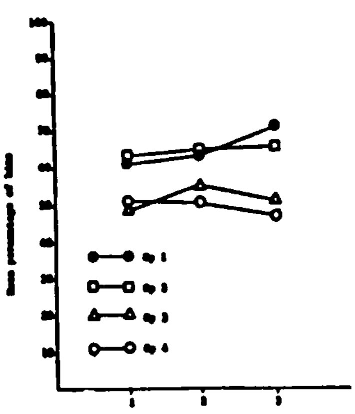


Figure 3. Mean percentage of hits per trial for Experiment 3

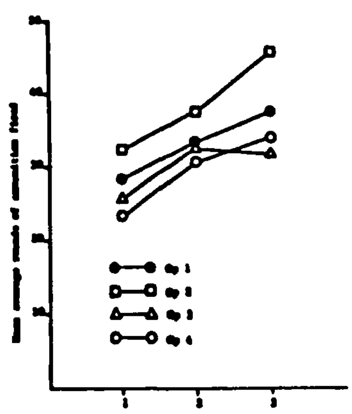


Figure 4. Mean average rounds of ammunition fired per trial for Experiment 3

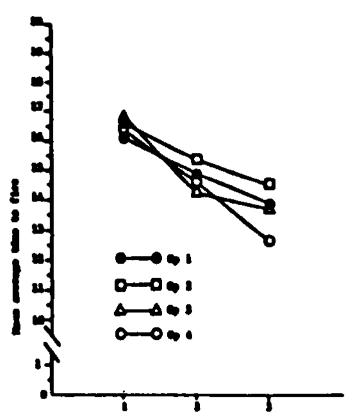


Figure 5. Mean average time to fire per trial for Experiment 3

The standard kill zone groups scored a significantly higher percent overall hits. This result indicates the dramatic effect on the subjects' accuracy when the target's vulnerable area is reduced. This could be accounted for by the characteristic prioritizing of speed over accuracy in video games in general. Subjects in all groups used typical video game strategy in destroying targets such as "fast shooting" without using appropriate tracking and laying of the gun on the targets. This can be supported by the significant decrease in the average time to fire over trials in all groups and the increase in number of rounds of ammunition fired. The subjects in the reduced kill zone groups continued to "shoot fast" across trials even though their percent hits suffered. Percent hits in the standard kill zone groups did not suffer from this "shoot fast" tactic because the larger kill zone allows for inappropriate tracking skills and less refined laying of the gun on the target. While there were no overall significant differences between the two game configurations on any of the measures, the nature of the learning curves for the groups did differ. The revised video game groups demonstrated a significant positive linear trend in percent hits across trials which indicates an increase in accuracy with practice. The standard video game groups did not significantly increase their hit percentage across trials. Both groups significantly decreased their average time to fire indicating that both game configuration groups became faster at firing rounds across game trials. The number of rounds fired also significantly increased across trials for all groups. Although all groups decreased their average time to fire and fired increasingly more rounds of ammunition, only the revised video game groups significantly increased their accuracy with practice. The standard video game groups did not demonstrate this improvement.

DISCUSSION

This research project was designed to address questions concerning the training value of the Battlesight tank gunnery game. Investigation into the training value of any device must necessarily begin with learning and whether learning actually occurs with practice. The results of both Experiments 1 and 2 indicate relatively strong practice effects. In Experiment 1, subjects continued to improve in accuracy across numerous practice trials when using the Gunner's primary sight. In Experiment 2, the learning curve indicates improvement in accuracy with practice when using the secondary sight on the Battlesight. The results of Experiment 3 suggest that the configuration of a video game could affect the game's optimal training value. A video game configuration which allows only one life and a limited amount of ammunition per game play with the ability to program a decreasing kill zone area to require greater accuracy for target kills may be beneficial. On the other hand, a standard video game format may be more highly motivating and yet not be appropriate in preserving positive training value.

Transfer of Training Between Two Tank Gunnery Training Devices

Bob G. Witmer
U.S. Army Research Institute-Fort Knox

Abstract

Twenty-four novice tank gunners were randomly assigned to two groups of 12 subjects each and trained on two gunnery simulators in succession. An M1 Conduct of Fire Trainer (COFT) and a Videodisk Gunnery Simulator (VIGS) were used. One group of subjects received COFT training first followed by VIGS training. The second group was trained on VIGS followed by COFT training. Improvement in gunnery performance was measured with each simulation device and training transfer between the devices was evaluated. Despite demonstrated increases in gunnery proficiency on each device and high similarity between the stimuli presented and responses performed on the devices, the results suggest that positive transfer does not occur. Results are discussed in terms of the specificity of skills learned on each device and the elements that must be included in training transfer research in order to properly evaluate simulators as training devices.

Introduction

Highly skilled operators can increase the effectiveness of modern weapon systems by engaging enemy targets with speed and accuracy. Speed and accuracy in tank gunnery is not easily acquired and requires considerable hands-on practice. Due to the excessive cost of ammunition, the number of rounds allocated for training has been cut in recent years. Required proficiency levels can no longer be achieved by training that relies solely on the expenditure of live ammunition. To close the training gap, tank gunnery training devices have been developed that permit extended firing practice with substantial cost savings. The present research investigates the training effectiveness of two of these devices, the MK1 Videodisk Gunnery Simulator (VIGS) and the M1 Unit Conduct of Fire Trainer (COFT). Each of these devices is designed to train and sustain M1 tank gunnery skills.

The M1 VIGS is a part-task, table-top trainer that utilizes video disk media to present filmed target scenes to the gunner. It is equipped with a single sight for engaging targets, and incorporates most of the switches and controls that the gunner operates during a live-fire engagement on the tank. The M1 COFT is a high fidelity whole-task gunnery trainer employing computer-generated target scenes and a highly realistic crew compartment. The crew compartment includes gunner's primary and secondary sights, the TC's sight and the TC's forward unity periscope. Nearly all of the switches and controls used by the gunner and TC in the M1 tank are simulated in the COFT.

The VIGS and the COFT are similar in that they both permit the gunner to acquire, track, and fire at moving and stationary targets. They differ in the number and types of engagements available and the sights and weapons simulated for firing those engagements. The COFT simulates more weapons and sights and has a much larger library of target engagements. The COFT also permits training the gunner and tank commander as a team whereas the VIGS does not. While the COFT has many more capabilities than the VIGS, it also costs about forty times as much. The VIGS used in this study presently sells for about $50,000, while the COFT costs about $2,000,000.

While the Army is proceeding to field the COFT and may field some form of the VIGS in the near future evidence is lacking to show that gunnery skills learned on the VIGS enhance performance on the COFT, or that skills learned on either the VIGS or COFT translate to increased proficiency on the tank. Several studies (Deason & Terrell, 1978, Kuma & McConville, 1982, Hoffman & Melching, 1983) have looked at transfer between VIGS or COFT and live fire gunnery performance, but none of these provide convincing evidence that transfer occurs. As Boldovici and Sabat (1985) state, "Nearly or perhaps all evaluations of Army training devices conducted to date have contained at least one design flaw of such severity as to preclude establishing a relation between device-mediated practice on one hand, and weapon system proficiency on the other." Some of the more common design flaws in recent evaluations are: (1) Small numbers of soldiers are used in the evaluation; (2) Subjects are not assigned randomly to groups; (3) Groups are treated differently in respects other than those under investigation; and (4) The criterion measure (e.g., live-fire gunnery) is not reliable.

The unreliability of live-fire tests suggests that alternatives are needed for evaluating tank gunnery training devices. Boldovici and Sabat (1985) suggest using dry fire and sight-picture photography, perhaps combined with device-based measures and verbal measures of critical knowledges as substitutes for live-fire tests. Another alternative is to determine if skills learned on one tank gunnery training device transfer and enhance performance on a second gunnery training device. Device-based performance measures are likely to be reliable, and positive transfer between devices would suggest that the devices are tapping common skills, perhaps the same skills that are required for skilled performance on the weapon system. The present research is designed to determine whether skills learned on the VIGS transfer to the COFT and vice versa.

Method

The subjects were 24 M60A3 armor crewmen. The average soldier tested had served in armor for 18 months. None of the subjects had more than one month of experience as a gunner and most were serving as drivers or loaders on M60A3 tanks.

Subjects were randomly assigned to two groups of 12 subjects each. One group of subjects, the VIGS-first group, was initially trained as gunners on the VIGS followed by testing on the COFT. The other group, the COFT-first group, was trained first on the COFT and then tested on the VIGS. Each group received the same number of trials (i.e., gunnery engagements) on both devices. The only difference between the groups was the order in which they

received training on the two devices. VIGS training consisted of engaging 30 single moving targets in two successive replications. COFT training was conducted in the same manner. Therefore subjects performed 60 engagements on the VIGS and 60 similar engagements on the COFT. All engagements were conducted in the thermal mode due to the unavailability of daylight target scenes for the prototype VIGS. Prior to training on either device, subjects were familiarized with the device and given four warm-up trials.

Measures of speed and accuracy were recorded for each engagement. Measures included opening time (time from target presentation until firing of the first round), hit time (time to achieve a target hit), and accuracy (hit or miss). Performance improvement was determined by comparing summary statistics for the first replication with statistics for the second replication using a correlated t-test. Training transfer from VIGS to the COFT was assessed by comparing the COFT performance of the VIGS-first group with that of the COFT-first group. Transfer from COFT to the VIGS was assessed by comparing the VIGS performance of the COFT-first group with VIGS performance of the VIGS-first group. Simple t-tests were used to make the comparisons.

Results

In evaluating transfer from one device to another, it is important to demonstrate that learning has occurred with the first device. The present research assesses learning by comparing performance during the second replication for each device with performance during the first replication. The results show significant improvements in VIGS performance from the first to the second replication for the three of the four dependent measures. Significant decreases in opening and hit times ($p<.01$) and a significant increase in hit percentage ($p<.025$) were obtained. Similarly COFT performance improves over replications. Significant decreases in opening and hit times ($p<.005$) and significant increases in percent first-round hits and overall hit percentage ($p<.001$) were found. These results suggest that subjects are learning something on each device that translates to improved gunnery performance on that device.

The question that remains is whether the skills learned on the first device transfer to the second. One approach to evaluating transfer is to measure subjects' performance on each device and then correlate the two sets of scores. Correlating VIGS and COFT scores yields significant positive correlations for hit percentage ($r=.56$, $N=24$, $p<.01$) and first-round hit percentage ($r=.58$, $N=24$, $p<.01$), but small nonsignificant correlations for opening and hit times. From these results, one might conclude (perhaps erroneously) that firing accuracy skills transfer from one device to the other. Gagne (1954) noted that high positive correlations between two performances do not establish the causal link which is necessary for demonstrating transfer between learning a skill on one device and learning or performing the skill on a second device. Such correlations simply suggest that similar skills or abilities are necessary for performing well on the two devices.

To establish training transfer, a more direct approach is required. In this experiment transfer from the VIGS to the COFT and from the COFT to the VIGS is assessed by comparing group differences using t-tests. Positive transfer from the VIGS to the COFT is indicated if the VIGS-first group

performs significantly better on the COFT than the COFT-first group. If the COFT-first group outperforms the VIGS-first group on the COFT, then negative transfer is indicated. Table 1 shows that skills learned on the VIGS do not transfer to the COFT, since none of the between group differences are significant.

Table 2 shows the results for transfer from the COFT to the VIGS. Superior VIGS performance by the COFT-first group would suggest positive transfer. But better performance by the VIGS-first group would indicate negative transfer. The results suggest that VIGS firing accuracy does not improve as a function of previous COFT training. However significant negative transfer for engagement speed is evident. That is, after training on the COFT, subjects exhibit significantly slower opening and hit times on the VIGS.

Table 1
VIGS to COFT Transfer of Training

COFT PERFORMANCE MEASURE		GROUP COFT FIRST	GROUP VIGS FIRST	MEAN DIFFERENCE
OPENING TIME	$\bar{X}$	16.9	16.0	-0.9 .N.S
	SD	(2.4)	(1.9)	
HIT TIME	$\bar{X}$	18.6	18.1	-0.5 .N.S
	SD	(3.2)	(2.3)	
% 1ST RND	$\bar{X}$	37.2	37.3	+0.1 .N.S
	SD	(13.6)	(15.2)	
% HITS	$\bar{X}$	39.1	40.7	+1.6 .N.S
	SD	(16.8)	(15.2)	

Table 2
COFT to VIGS Transfer of Training

VIGS PERFORMANCE MEASURE		GROUP VIGS FIRST	GROUP COFT FIRST	MEAN DIFFERENCE
OPENING TIME	$\bar{X}$	7.8	9.8	+2.0,P<.005
	SD	(1.3)	(2.4)	
HIT TIME	$\bar{X}$	8.9	11.0	+2.1,P<.0005
	SD	(1.2)	(2.2)	
% 1ST RND	$\bar{X}$	66.9	63.1	-3.8, N.S
	SD	(12.2)	(9.4)	
% HITS	$\bar{X}$	79.5	78.1	-1.4, N.S
	SD	(8.3)	(9.0)	

Discussion

The results described are not what one might have expected based on the similarities between the gunnery tasks performed on the two devices. While significant positive correlations were obtained between accuracy measures on the two devices, no evidence of positive transfer for accuracy was obtained. Furthermore, negative transfer from the COFT to the VIGS was obtained for

opening and fire times. The slower times on the VIGS for the COFT-first group are likely due to habits formed on the COFT that were inappropriate to performing on the VIGS. For example, on the COFT the gunner must wait for the TC to perform several actions before engaging the target. Although the gunner need not wait on the VIGS, the habit of waiting may have carried over to influence VIGS performance. The COFT also requires the gunner to aim more carefully than the VIGS to "kill" the target. On the COFT the gunner must place the reticle within .67 mils of the center of mass to get a target kill, whereas the VIGS requires only that the target be hit. Aiming more accurately requires additional time. Thus the COFT-first group may have continued to aim carefully when they engaged targets on the VIGS, and consequently took more time to open fire and hit targets than the VIGS-first group.

The lack of positive transfer between the VIGS and COFT for the accuracy measures is more puzzling. Differences in the response and performance of the gunner's control handles on the two devices may explain the results. Learning to operate the control handles to track and lay the reticle on target may require finer movements on one device and the proprioceptive feedback provided in operating the handles may differ for the two devices. This may reduce or eliminate transfer of tracking and aiming skills between the VIGS and COFT. An alternate explanation for the lack of positive transfer is that the training period with each device was too brief to produce transfer. While the demonstration of within device performance improvement weakens this argument, further research is needed to determine if longer or repeated periods of practice with one device results in positive transfer to the other device. Currently, it appears that skills learned on the VIGS do not transfer to the COFT and vice versa. Thus those individuals responsible for training gunnery skills should not expect that skills learned during a short period of practice with the VIGS or the COFT will transfer to the other device. The present study provides no support for a training strategy where basic gunnery skills are learned on the VIGS and these skills are subsequently enhanced and broadened on the COFT.

REFERENCES

Boldovici, J.A. and Sabat, S.R. Measuring transfer from training devices to weapon system. Unpublished manuscript, 1985.

Deason, P., Robinson, M. and Terrell, G. (1978). Guns over Boise, Snake River Shoot-out phase. White Sands Missle Range, NM: US Army TRADOC Systems Analysis Activity (TRASANA).

Gagne, R.M. (1954). Training devices and simulators: Some research issues. American Psychologist, 9(3).

Hoffman, R.G. and Melching, W.H. (1983). Field trials of the MK60 tank gunnery simulator in armor institutional training courses, Vol I. (HumRRO Final Report FR-TRD(KY)-82-9). Alexandria, VA: Human Resources Research Organization.

Kuma, D. and McConville, L. (1982). Independent evaluation report for M1/M60 series Unit Conduct of Fire Trainer (UCOFT). Fort Knox, KY: US Army Armor Center.

Training Development for Vehicle Identification:
Special Problems Related to Infrared Systems

Nigel R. Nicholson and Norman D. Smith
Army Research Institute, Fort Hood Field Unit

ABSTRACT

Infrared (IR) technology is used on increasing numbers of military systems. Two prominent examples are the M1 Abrams tank and the M60A3 tank One of the primary objectives of the IR is to provide visual information to the soldier, be he tank commander, gunner, or pilot; about targets he may be expected to destroy. This paper will discuss the Target Acquisition and Analysis Training System research program at ARI and the development, testing, and subsequent Army-wide distribution of a training program in the use of the IR system on the M1 and M60 tanks for target recognition and identification.

INTRODUCTION

Since the introduction of large-scale mechanized warfare in World War II, the armed services have experienced an increasing need to develop skills in target acquisition. A target is acquired when its presence has been detected, recognized, and identified sufficiently well to permit engagement. Yet, because modern technology has produced weapons and fire-control systems which can engage targets at ranges far in excess of that possible with the naked eye, target acquisition has come to rely increasingly on electronic aids. In 1980, the Army's proponent for vehicle recognition and identification (R&I) at the Combined Arms Center (CAC), Fort Leavenworth, Kansas tasked the U.S. Army Research Institute to produce a series of training programs in R&I. Thus, the Target Acquisition and Analysis Training System (TAATS) research at Fort Hood Field Unit began. The first product from TAATS was the Basic Combat Vehicle Identification (CVI) training program developed in 1981 and adopted by the Army that same year as its standard R&I training program, GTA 17-2-9. For details see Smith, Heuckeroth, Warnick and Essig; 1980.

Thermal Simulation Development

In November 1982, CAC placed highest priority on the development and evaluation of a thermal (or infrared) combat vehicle identification training program (TCVI) in response to urgent requests from commanders in Europe and from the Deputy Chief of Staff, Training at TRADOC, because the infrared sight on the M1 tank had been developed without concommittant training in its use.

In order to develop an IR combat vehicle identification program it was necessary to locate thermal images of the 30 vehicles in the Basic CVI Program. Assistance was solicited from a variety of in-house laboratories including Nightvision and Electro-optical Laboratory (NV&EOL) as well as through Electro-optical and Infrared Measurements (E-OIR), an ARI contractor. Although the search was thorough, few of the images obtained met the specifications necessary to design a training program. In part, the failure to meet training needs resulted from the fact that a large percentage of the

images were taken with the settings on the tank sights adjusted so that any part of the vehicles that radiated heat showed up as white. The white-hot image, although extremely good for detection, becomes a blurr or blob as range increases. Beyond 1,400 meters the image becomes progressively degraded and essentially useless for R&I.

So that a solution to the training challenge presented by the white-hot image could be found, over 3,000 photographs of a limited array of vehicles were taken through the M60A3 and M1 tank IR sights. By changing the polarity switch on the sight to black-hot, the image took on the characteristics of a silhouette that maintained its shape much better than did the white-hot image, particularly at longer ranges. This finding minimized the problem of image instability.

One major concern remained: where to acquire black-hot images of the 30 vehicles in the CVI training program photographed through an M1 Tank Integrated Sight or M60A3 Tank Thermal Sight at the specified ranges needed for training? This problem seemed insurmountable given the short time available to produce the TCVI program unless a means of simulating the maximum-high-contrast image in the black-hot setting were possible. Experimentation resulted in images that were markedly similar to the actual images obtained through the TTS and TIS sights. Based on the knowledge of how to simulate those images, simulations were made of the remaining vehicles in the training program.

Thermal Simulation Procedure

A terrain background drawn on paper was placed on a table constructed of matted acrylic plastic shaped to provide an even transition from background to foreground. A reference point was marked on background and each model was placed, in turn, on the reference point. Front and rear lighting were balanced to provide a silhouette image. A video camera was focused on the models and the video signal was fed to a high-resolution (600-line) monitor. Image size was adjusted to provide compatibility with the CVI sizing template. Photographs were taken of the face of the monitor with a Nikon 35mm camera mounted on a tripod. The film used was Kodak 135-36 Ektachrome 64 professional.

METHOD

Soldiers Trained

The data for this research were gathered from 41 soldiers who were assigned to the 1/68 Armor Battalion stationed at Wildflecken, Federal Republic of Germany.

Training Procedure

The soldiers were administered a CVI/TCVI test before and after training. The test was comprised of 60 photopic slides (i.e. actual photographs) and 60 simulated thermal images. All soldiers at received only TCVI training and CVI/TCVI pre and posttraining testing.

Program Description

The TCVI training program consists of six training modules and a final test module. Each training module has four views of five separate vehicles. During the training modules the IR images are presented in combination with comparable views of the photopic image of each vehicle from the Basic CVI Training Program. Each training module is shown twice, once with instructor-determined exposure and a second time for 15 seconds. Each module also has a test comprised of simulated IR images of the five vehicles with each shown for 8 seconds. The final test of the 30 vehicles, which displays the IR image for 8 seconds, provides an overall measure of training progress. An instructor's manual which contains all the information needed to recognize and identify the vehicle accompanies each module. As in the Basic CVI Program, only key cues are addressed. For a complete description of the TCVI Program, see GTA 17-2-10 available from TASO.

In the evaluation research of the TCVI program (Smith, Shope & Heuckeroth, 1984) a modification was made in the test module in the TCVI program to permit comparison of performance changes resulting from training on the photopic as well as the thermal images. Presentation of the photopic and thermal slides was randomized during the pretest and posttest.

Research Objectives

The research was designed to evaluate the effectiveness of the Basic TCVI program on recognition and identification performance.

Data Collection Instruments

The soldiers were given a prepared answer sheet on which to record their responses. They had first to make a recognition response, i.e., categorize the vehicles as a Friend (F) or Threat (T); then identify the vehicle by name or number. For example, if a Soviet T-62 was projected, the soldier should recognize it with "T" for threat and identify it with "T-62."

RESULTS

Performance Change Resulting from TCVI Training

Analyses of variance were performed on the pretraining and posttraining scores that were obtained from the TCVI data. The effects of several variables were examined in the ANOVA and are presented in Smith et al. (1984).

However, because the present paper will discuss only the effect of training, an extract from the ANOVA reported by Smith, et al. (1984) is presented. As a result of TCVI training, a significant improvement was found for both recognition ($F(1,40) = 61.67$, $p < .001$), and identification ($F(1,40) = 108.62$, $p < .001$). The means and standard deviations are presented in Table 1.

Table 1

Comparison of Pretraining and Posttraining Recognition and Identification Performance Scores[a] (n=41)

	Recognition		Identification	
	M	SD	M	SD
Pretraining	81.90	1.46	18.61	6.68
Posttraining	97.10	13.78	40.69	17.68

[a] Maximum possible score is 120. The mean scores display the mean number of correct responses to 60 photopic and 60 thermal images presented in a randomized order within each test session.

GT Scores and Performance

An analysis of variance revealed significance for GT score in both the pretest and the posttest for recognition ($F(2,25) = 3.30$, $p < .05$); and for identification ($F(2,25) = 8.46$, $p < .002$). A Duncan Multiple Range Test of the identification means for the posttest indicated that soldiers with GT scores above 100 performed significantly better ($p < .05$) than did the others.

DISCUSSION AND CONCLUSIONS

Discussion

After reviewing the available information, the ARI technical support team from HumRRO experimented photographically with images through IR sights and found that the black-hot setting mitigated the deficiencies of the white-hot image. Only then was it feasible to conclude that a long range target identification training program could be produced.

A significant increase in the number of vehicles identified was found after training. A comparison with other programs was not possible because the TCVI was the only available IR program. The TCVI shares similar objectives with the other CVI programs: to keep training simple with a minimum of support materials; to train soldiers to recognize primarily those cues that discriminate vehicles at realistic combat ranges; and to be modular in design for use in short training periods. The TCVI achieved these objectives.

An interesting finding is that performance decreases significantly as GT scores fall below 100. This effect is more apparent with identification than recognition because the probability of being right by chance alone on recognition is .5 but only .03 in identification. Earlier studies (Smith et al. 1985) have generally found GT scores to correlate with CVI performance

scores within the r = .75 - .82 range. More interestingly, however, the most powerful predicator of subsequent CVI performance scores has been the soldiers' initial CVI performance score: approximately one-third of the soldiers who scored low on the first posttraining test continue to perform poorly up to three training sessions later. There is clearly a need for continuing research to uncover the responsible variables. Whatever the cause of low CVI performance scores the performance decrement has implications for training and research. First, it may be necessary to make provision for personnel below the GT level of 100 to receive more CVI training. Second, additional research is needed to ascertain that poor motivation does not account for part of the decreased performance. A third consideration is the relationship between GT level and long-term memory, with specific attention to the frequency with which training must be repeated.

It is highly desirable to determine the degree to which generalization in learning takes place from any simulation training to the real situation. However, the cost of conducting a field test involving the 30 CVI vehicles might far exceed the probable value of knowing whether 70%, 80%, or 100% of what is learned transfers to the field. Current research projects are pursuing the overall generalization question. At present, a good review of the literature on generalization of learning from simulation training is found in Baum et. al. (1982). Hence, until resources are made available, the estimate of training generalization must rest primarily on face validity.

General Conclusions

The findings of this research led to the following general conclusions:

- The Basic Thermal Combat Vehicle Identification (TCVI) Training Program substantially improved performance compared to pretraining performance.

- Performance on the TCVI Program varies widely. Consequently, the Army should consider providing additional training sessions for soldiers whose performance scores are in the lower half of the distribution.

REFERENCES

Baum, D.R., Smith, D.A., Hirschfeld, C.A., Klein, C.A., Swezey, R.W., & Harp, R.T. (1982). Specification of training simulator fidelity: A research plan (Technical Report 558). U.S. Army Research Institute for the Behavioral and Social Sciences.

Foskett, R.J., Baldwin, R.D., & Kubala, A.L. (1978). The detection ranges of features of armored vehicles (Technical Report TR-78-A37). U.S. Army Research Institute for the Behavioral and Social Sciences.

Ratches, J. & Swistak, J. (1972). Thermal systems' performance evaluation using photographic thermal imagery (Technical Report 7014). U.S. Army Electronics Command.

Smith, N.D., Heuckeroth, O.H., Shope, G.L., Warnick, W.L., & Essig, S.S. (1984). Target acquisition and analysis training system: Development and Evaluation of the Basic Thermal Combat Vehicle Identification Training Program (Research Report 1378). U.S. Army Research Institute for the Behavioral and Social Sciences.

Smith, N.D., Heuckeroth, O.H., Warnick, W.L., & Essig, S.S. (1980). Evaluation of a new approach to target acquisition training: The Combat Vehicle Identification (CVI) Training Program (Research Report 1304). U.S. Army Research Institute for the Behavioral and Social Sciences.

Smith, N.D., Heuckeroth, O.H., Shope, G.L., Warnick, W.L., and Essig, S.S. (1985). Target acquisition and analysis training system: Effects on combat vehicle identification (CVI) performance of the number of vehicles presented and the frequency of training sessions.

THE EFFECTS OF SIMULATOR DELAYS ON THE ACQUISITION OF FLIGHT CONTROL SKILLS

TERENCE J. FISCHER
SYNERGY INC.

GARY E. RICCIO
SYSTEMS RESEARCH LABORATORIES INC.

GRANT R. McMILLAN
ARMSTRONG AEROSPACE MEDICAL RESEARCH LABORATORY

ABSTRACT

This report describes the results of four experiments investigating the effects of transport delays on the control of altitude and roll in a simulated F-16. The transport delays in the visual display ranged from approximately 20 ms. to 340 ms. The results show that delays affect both workload and flight control. The effects of delay on the transfer of training were complex and not necessarily predictable from the effects during training.

INTRODUCTION

Simulators as training devices are being used more than ever before. With new display technologies the ability to reproduce visual scenes has become a much easier task. The current belief in simulator design is that a more complex display will produce a better simulation. On the surface this seems to be an accurate assumption. The generation of complex displays however can lead to problems.

When an input is delivered to the controls of the simulator a computer processing action takes place. This processing is necessary for the simulator to alter its display in response to the given input. As the simulator designer strives for a greater correspondence between the simulator display and the actual scenery being simulated, processing time will increase. This is due to increased computational load. Consquently the simulator time lag is greater than the time lag found in actual aircraft.

One use of a simulator is in pre-flight training. A problem may occur with a simulator used for training if the inherent time delay (TD) in the system detracts from the fidelity of the simulation. Previous studies have shown that TDs can affect pilot performance (Ricard and Puig, 1977). However there are no data on the effects of TD on transfer of training (TOT).

This report describes the results from four experiments on the effects of simulator delays. Separate experiments examined the effects of TD on roll and altitude control. Two experiments provide information about the effects of TD on flight control performance and workload. The other two experiments provide information about TOT from simulators with long TDs to a simulator with a short TD.

METHODS

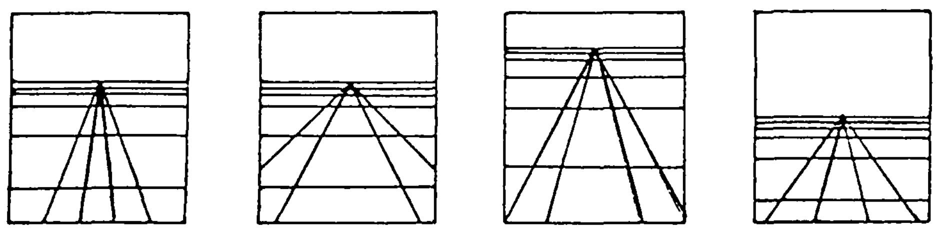
<u>Subjects</u> Twenty professional subjects that were not experienced in this type of flight task were used. None of the subjects were pilots. The subjects ages ranged from eighteen to thirty-four.

<u>Apparatus</u> For the altitude control experiments the flight dynamics were simulated using an analog computer (EAI 2000). The roll experiments used a digital simulation implemented on a PDP 11/60 computer. Flight dynamics were those of an F-16 flying 400 knots at 100 ft. The inputs to the computers included the operator's control inputs and the simulated wind gust. Outputs were the states of the aircraft. The visual display was then generated with a

DEC GT-40 (vector graphics) display processor and presented on a Hewlett-Packard high-speed phosphor monitor. The screen was 38 cm wide and 28 cm high. The corners of the screen were masked with rounded nonreflective black cardboard to eliminate any extraneous cues. Viewing the display from approximately 38 cm resulted in a 53 deg wide field of view.

The cockpit consisted of the display screen and a seat with a force stick mounted on the right hand side. The seat was similar to those found in high performance aircraft. The cockpit was enclosed in a room with no internal light sources and nonreflective black walls.

Display The display scene consisted of a grid pattern giving a perspective view of flat terrain and the horizon. When the aircraft was level the horizon was at eye level. In the altitude control experiments the perspective angle changed as a function of altitude and the position of the horizon changed as a function of pitch (see Figure 1). Altitude was continuously perturbed by a simulated wind disturbance that was uncorrelated with the vehicle states. This quasi-random forcing function was generated by a sum of thirteen harmonically unrelated sinusoids (Levison, Zacharias, and Sinacori, 1982). Pitch and altitude were also affected by the forces applied to the control stick.



Level Pitch High Altitude — Level Pitch Low Altitude — Pitch Down — Pitch Up

Figure 1

For the roll control experiments, the display was identical to that used in the altitude control experiment. The perspective angle and the horizon line height however remained constant. The scenery revolved about the vanishing point for the longitudinal components of the terrain grid. The roll attitude was continuously perturbed with a sum-of-sines forcing function. The spectral power distribution for this forcing function was similar to that used in the altitude control experiments.

Initial Conditions Four assigned altitudes were used in the altitude control experiments (50, 100, 150 and 200 ft.). The subject's task was to maintain the assigned altitude. At the beginning of each trial the subjects were allowed to memorize the current altitude. During the trial there were no reference points for the subjects to use, only their memory of the display at the assigned altitude. In the roll control experiments the subjects were also given no reference points. The altitude in the roll control experiments (100 ft.) remained constant across trials and subjects.

Time Delays The steady state TDs between the subject's control actions and the display response in the altitude control experiments were 20, 40, 80, 160, and 320 ms. The TDs used in the roll control experiments were 43, 63, 103, 183, and 343 ms. The reported TDs do not include the phase lag of the simulated aircraft dynamics.

Procedure The subject's task was to either maintain a constant altitude or a wings-level attitude condition. Each flight began with 15 to 30 seconds in which the subject could memorize the characteristics of the initial condition. A ready signal was given and the trial began when the gust and the control stick were activated. The trial lasted 113 sec of which the last 81.92 sec were used for analysis. The subjects gave their subjective rating of the workload at the end of each trial. The subject's mean, standard deviation, and RMS roll or height error were displayed. A session included four flights and subjects ran two sessions per day. Subjects also engaged in a concurrent interval production task (IPT) on some trials. Variability of intervals in the IPT was used as a workload measure (Shingledecker, 1983). This required that the subject tap at a constant rate (2 taps per sec.) with their left thumb while controlling the aircraft with their right hand. Before the first trial in each session the subject performed the IPT alone. In the first and second trials the subject performed the flight task alone. In the third and fourth trials the subject performed both the flight task and the IPT.

Experiment 1: This was an altitude control experiment using a between subjects design. Two subjects were assigned to each of five delay conditions. The subjects performed only the flight control task (IPT and workload ratings were not employed). The subjects were trained for ten sessions. They then transferred to the minimal delay condition (20 ms) for ten transfer sessions. Subjects were given the same assigned altitude (100 ft) for all flights in the first two sessions of the training and transfer phases of the experiment. Each of the four altitudes was assigned in all other sessions. Order of presentation within each session was counterbalanced across sessions by the Latin square method.

Experiment 2: This was a within subjects altitude control experiment using five subjects from experiment 1 (one subject from each training condition). Each subject ran four sessions in each TD for a total of twenty sessions. Each of the four altitudes was assigned in all sessions. Order of presentation within each session was counterbalanced across sessions by the Latin square method.

Experiment 3: This was a roll control experiment using a between subjects design. Two subjects were assigned to each of five delay conditions. The subjects were trained for ten sessions. They then transferred to the minimal delay condition (43 ms) for ten transfer sessions.

Experiment 4: This was a within subjects roll control experiment using five subjects from experiment 3 (one subject from each training condition). Each subject ran four sessions in each TD for a total of twenty sessions.

RESULTS

Analysis A logarithmic transformation was applied to the flight control data (RMS error) to produce greater homogeneity of variance and to increase the linear correlations between the dependent and independent variables. General linear model analyses were performed on the transformed data using the Statistical Analysis System (SAS) on a VAX 780. In the between subjects experiments TD was coded as a classification variable. TD was coded as a continuous variable in the within subjects experiments.

Experiment 1: The results for the altitude control task are depicted in Figure 2. The effect of TD was not significant in the training phase of the experiment. In the transfer phase of the experiment the group trained with the 160 ms simulation obtained the best scores (lowest RMS error) followed by the 80, 40, 20, and 320 ms groups. Bonferroni t tests showed that the only significant differences were between the 160 ms group and both the 20 and 320 ms groups.

EXP #1 PITCH CONTROL

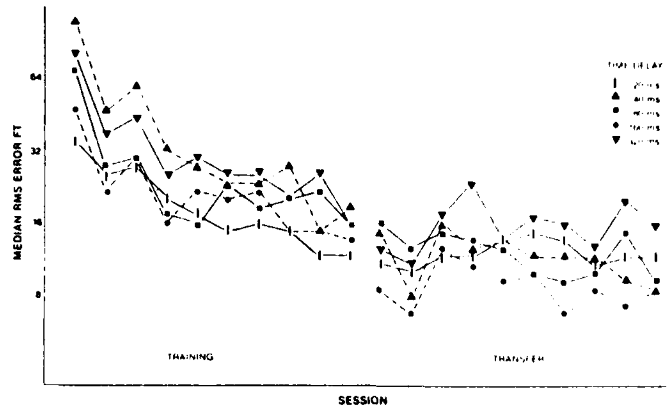


Figure 2

Experiment 2: There was a significant affect of TD on the RMS altitude error ($R=.42875$, $F(1,381)=74.75$, $p<.0001$), subjective workload ($R=0.11949$, $F(1,381)=7.73$, $p<.0057$), and IPT ($R=0.11389$, $F(1,180)=6.41$, $p<.0122$). A canonical discriminant analysis showed that RMS altitude error made the major contribution to the linear function (of the dependent variables) that best discriminated among the TD conditions. The total structure coefficients for RMS error, subjective workload, and IPT were 0.91, 0.24, 0.27, respectively. The canonical correlation between the linear discriminant function and TD was 0.45.

Experiment 3: The results for the roll control task are depicted in Figure 3. Bonferroni t tests for the training phase of the experiment showed that the RMS roll error obtained with the 343 ms delay was significantly different from all other TDs. There were no significant differences among the other four TDs. The transfer phase of the experiment revealed performance across TDs that was similar to that in experiment 1. The best scores were obtained by the group trained with the 183 ms TD, followed by the 63, 103, 43, and 343 ms TD groups. Bonferroni t tests showed that the only significant differences were between the 343 ms group and the 103, 63, and 183 ms groups. The smallest TD was not significantly different from the longest TD.

Experiment 4: There was a significant effect of TD on RMS roll error ($R=0.90396$, $F(1,390)=3190.06$, $p<.0001$) and upon the subjective workload ($R=0.43971$, $F(1,390)=167.07$, $p<.0001$). The canonical discriminant analysis showed that RMS roll error made the major contribution to the linear function (of the dependent variables) that best discriminated among the delay conditions. The total structure coefficients for RMS error, subjective workload, and IPT were 0.98, 0.38, and 0.01, respectively. The canonical correlation between the linear discriminant function and the TD was 0.92.

DISCUSSION

The within subjects experiments provide the most sensitive tests of the effects of simulator delays on flight control performance and workload. While there are significant effects of TD on RMS error and subjective workload in both the altitude and roll control experiments, the effects are somewhat larger for roll control. This is consistent with results from other laboratories (Ricard and Puig, 1977). Canonical discriminant analysis shows that TD affects flight control performance more than workload. This indicates that the effect of TD on the flight task cannot be ameliorated by working harder.

EXP #3 ROLL CONTROL

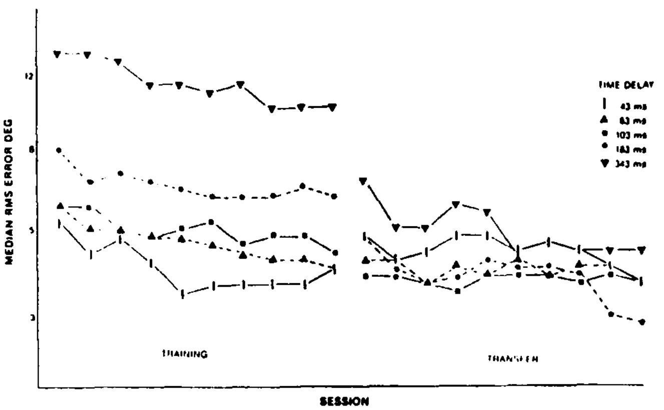


Figure 3

The between subjects experiments provide the first data on the effects of simulator delays on TOT. This is important because TDs in training simulators are generally much larger than those in real aircraft. Both the altitude and roll control experiments indicate that training with long TDs (greater than 0.3 s) can be significantly worse than training with moderate TDs (around 0.1 s). Interestingly, these experiments also suggest that moderate TDs may actually have a facilitating effect on training. Moderate TDs may force subjects to respond more quickly to maintain adequate flight control in demanding situations. However, long TDs can create instabilities such as the well known pilot induced oscillation.

This investigation is considered to be exploratory given the small number of subjects and the limited degrees of freedom in the flight control tasks. The present investigation suggests that experiments should include a set of TDs that could reveal nonmonotonic effects on TOT. Current experiments are examining the effects of TD on transfer of training in six degree-of-freedom simulations.

The experiments presented here are part of an ongoing investigation into simulator TDs. Some of the other problems to be investigated are: (1) The tradeoff between display complexity and computational TDs, and (2) The effects of temporal asynchrony or mismatches in multimodal simulations.

References

Levison, W.H., Zacharias, G.L., and Sinacori, J.B. (1982). Design of an Experiment to Study the Pilot's Use of Visual and Motion Cues in a Height-Regulation Task. (BBN Report No. 5028). Cambridge, MA: Bolt, Beranek, and Newman, Inc.

Ricard, G.L., and Puig, J.A. (1977). Delay of Visual Feedback in Aircraft Simulators (NAVTRAEQUIPCEN TN-56). Orlando, FL: Naval Training Equipment Center

Shingledecker, C.A., Acton, W.H., and Crabtree, M.S. (1983). Development and Application of a Criterion Task Set for Workload Metric Evaluation. Proceedings of the Second Aerospace Behavioral Engineering Technology Conference. Society of Automotive Engineers.

ADEQUACY OF THE TIME ALLOCATED FOR MEETING ARMY NATIONAL GUARD AVIATOR TRAINING REQUIREMENTS

Sandra M. Szabo, John W. Ruffner,
Kenneth D. Cross, and Michael G. Sanders

ABSTRACT

Army National Guard (ARNG) aviators must meet the same aviation training requirements as active Army aviators. During the past 10 years, the training requirements have significantly increased; yet, the amount of time allocated for ARNG aviators to meet the requirements has remained relatively constant. To determine if the aviators need additional allocated training time, a questionnaire survey was conducted. The survey indicates that ARNG aviators perceive the time allocated for meeting their current training requirements to be generally inadequate. The major obstacles to meeting the requirements are an Insufficient Number of Flight Hours, Unavailability of Instructor Pilots, Unavailability of Training Support Areas, Unavailability of Support Equipment, and an Insufficient Amount of Personal Time. The data suggest that, although additional time may be needed to meet the training requirements, remedial actions other than an increase in allocated training time can also be taken.

INTRODUCTION

An aviator in the Army National Guard (ARNG) must meet the same aviation training requirements as an aviator in the active Army. Due to modernization of the Army's aviation fleet during the past ten years, the operational requirements of the aircraft have significantly increased; consequently, aviators in the current force must acquire and maintain additional and increasingly complex aviation skills.

Despite the increase in the number and complexity of the training requirements, the amount of time allocated for ARNG aviation training has remained relatively constant since the 1970s. Generally speaking, ARNG aviators must meet their annual training requirements during a combination of 48 Unit Training Assemblies (UTAs), 24 Additional Flight Training Periods (AFTPs), and 15 Annual Training (AT) Days. The problem of meeting the requirements in a limited amount of training time is exacerbated by additional factors, such as the aviators' commitments to their civilian jobs and the geographical distances between the aviators' homes and the aviation training facilities.

Recognizing that ARNG aviators may not be able to meet existing training requirements in the amount of time that is currently allocated, the National Guard Bureau (NGB) requested that the Army Research Institute (ARI) Field Unit at Fort Rucker, Alabama, conduct research to determine if additional training time is needed. The results described in this report are derived from Phase I of the research.

METHODOLOGY

Phase I consists of a questionnaire survey of all ARNG aviators. The questionnaires were administered to the aviators during a weekend drill period. Over 3,600 (77%) of the aviators in the ARNG completed the questionnaire.

In one part of the questionnaire, the aviators used 7-point rating scales to indicate (a) the perceived adequacy of the training requirements for maintaining a safe level of aviation proficiency, and (b) the perceived adequacy of the time allocated to meet the training requirements. The scale for rating the adequacy of the requirements ranged from "1," indicating that the requirement is "Much Less than Adequate" to "7," indicating that the requirement is "Much More than Adequate." A rating of "4" indicated that the requirement is "About Right." The scale for rating the adequacy of the training time ranged from "1," indicating "Too Little Time Allocated to the Requirement" to "7," indicating "Too Much Time Allocated to the Requirement." A rating of "4" indicated that the time is "About Right." The aviators provided the ratings for each of the following specific continuation training requirements:

- Emergency Tasks,
- Emergency Procedures,
- Instrument Tasks,
- Terrain Flight (NOE) Tasks,
- Unaided Night Tactical Tasks,
- Night Vision Goggle (NVG) Tasks,
- Tactical/Special Tasks,
- Mission Tasks, and
- Additional Tasks.

In another part of the questionnaire, the aviators checked each of a number of factors that pose obstacles to meeting the continuation training requirements. The obstacles include:

- Unavailability of Instructor Pilots (IPs),
- Unavailability of Support Personnel,
- Unavailability of Aircraft,
- Unavailability of Support Equipment,
- Unsatisfactory Operational Hours of the Army Aviation Support Facility,
- Unavailability of Training Support Areas,
- Insufficient Number of Flight Hours,
- Nonaviation Obstacles, and
- Insufficient Amount of Personal Time.

RESULTS

The mean ratings of the adequacy of the training requirements for maintaining a safe level of aviation proficiency and the adequacy of the time allocated for meeting the requirements are shown in Figure 1. Because of the large sample sizes for each requirement, statistical analyses (Cohen, 1977) were conducted to determine both the statistical and the practical significance of the difference between each observed mean and the hypothesized mean of "4." The analyses identify the practical differences as small, medium, or large effect sizes; medium and large effect sizes were operationally defined as having practical significance.

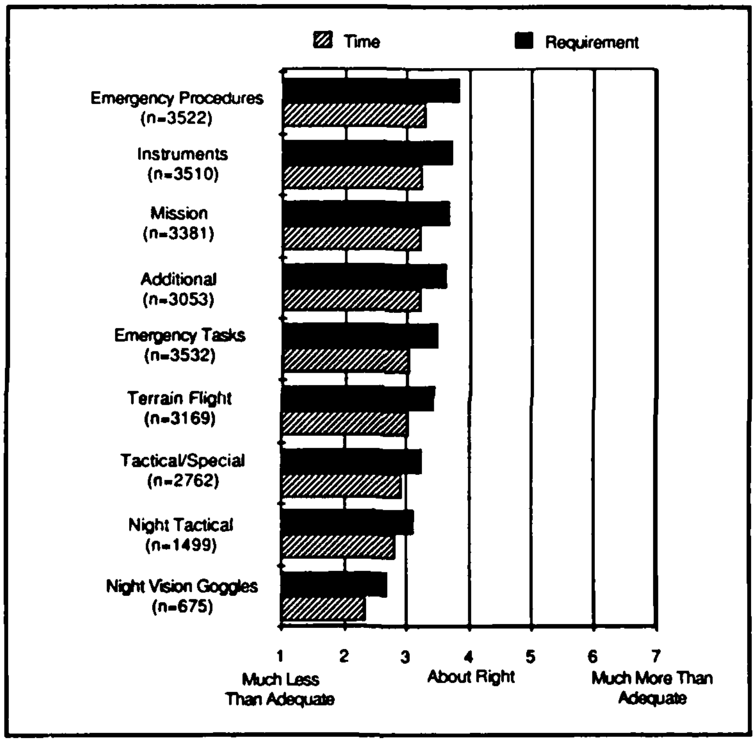


Figure 1. Mean ratings of the adequacy of the requirements and time: Continuation Training Requirements.

The results of the analyses of the adequacy of the training requirements indicate that the mean rating of the adequacy of Night Vision Goggle (NVG) training requirements is sufficiently less than "4" to represent a large effect size. The mean ratings of the adequacy of Unaided Night Tactical and Tactical/Special flight requirements are sufficiently less than "4" to represent medium effect sizes. The differences between the observed and hypothesized mean ratings for all the remaining requirements, except Emergency Procedures, represent small

effect sizes. The results of the analyses of the adequacy of training time indicate that the mean ratings of the adequacy of time for meeting NVG, Unaided Night Tactical, and Tactical/Special tasks are sufficiently less than "4" to represent large effect sizes. The differences between the observed and hypothesized means for the remaining requirements represent medium effect sizes.

The rating data suggest that ARNG aviators are unable to meet their training requirements in the amount of time that is currently allocated; therefore, an increase in the total amount of training time seems necessary. It is possible, however, that the amount of time is adequate; yet, certain factors interfere with the utilization of the time. To provide information about this possibility, the percentage of aviators identifying each obstacle for each continuation training requirement was computed. The percentages and the number of respondents on which the percentages are based are summarized in Table 1. Percentages that equal or exceed 25% are considered to be operationally significant and are noted by an asterisk.

TABLE 1

PERCENTAGE OF AVIATORS IDENTIFYING OBSTACLES TO MEETING CONTINUATION TRAINING REQUIREMENTS

	OBSTACLE								
CONTINUATION TRAINING REQUIREMENT	INSTRUCTOR PILOT	SUPPORT PERSONNEL	AIRCRAFT	EQUIPMENT	AASF HOURS	TRAINING AREAS	FLIGHT HOURS	NON-AVIATION	PERSONAL TIME
Emergency Tasks (n=3376)	36*	04	19	08	16	09	32*	23	25*
Emergency Procedures (n=3344)	25*	03	08	12	11	11	16	19	26*
Instrument Tasks (n=3374)	27*	03	23	16	13	06	30*	20	24
Terrain Flight (n=3143)	21	05	14	08	11	40*	30*	20	22
Unaided Night Tactical (n=1737)	23	05	16	16	15	27*	31*	16	24
Night Vission Goggle (n=1190)	28*	06	19	44*	12	28*	29*	15	24
Tactical/Special (n=2785)	17	07	15	16	12	26*	31*	22	23
Mission (n=3238)	13	07	15	14	12	19	30*	24	23
Additional (n=3038)	11	05	12	10	10	12	26*	22	25*

Key: n = total number of aviators responding to each item.

NOTE: Obstacles considered by 25% or more of the aviators are identified by an astrisk (*).

The data presented in Table 1 indicate that the NVG training requirement, for which the training time is perceived as most inadequate, is affected by the greatest number of obstacles. Unavailability of Support Equipment (i.e., night vision goggles and cockpit lighting) represents the single greatest obstacle for meeting this requirement. Unavailability of IPs, Unavailability of Training Support Areas, and an Insufficient Number of Flight Hours also pose major obstacles to meeting NVG requirements.

An Insufficient Number of Flight Hours is a major obstacle for meeting all continuation training requirements except Emergency Procedures, which are often performed in the synthetic flight training system (SFTS). Unavailability of IPs and Unavailability of Training Support Areas (e.g., firing ranges, NOE sites) also pose significant problems for meeting a number of continuation training requirements. Specifically, Unavailability of IPs is a major problem for meeting all continuation training requirements that require the presence of an IP (i.e., Emergency Tasks, Emergency Procedures, Instrument Tasks, and NVG). Unavailability of Training Support Areas is a problem for meeting all tactical training requirements (i.e., Terrain Flight, Unaided Night Tactical, NVG, and Tactical/Special). Finally, an Insufficient Amount of Personal Time is a major obstacle for meeting the requirements for Emergency Tasks and Emergency Procedures.

CONCLUSIONS

The questionnaire data provide two sources of information for concluding that additional training time is needed to meet current ARNG aviation training requirements. First, the rating scale data indicate that ARNG aviators judge the training time to be inadequate for meeting all continuation training requirements. The time is particularly inadequate for meeting NVG, Unaided Night Tactical, and Tactical/Special flight requirements; furthermore, the aviators judge that these requirements are inadequate for maintaining a safe level of aviation proficiency. Second, the checklist data indicate that an Insufficient Number of Flight Hours poses the single greatest obstacle for meeting all continuation training requirements. This finding further supports the need for additional training time. However, the identification of non-time-related obstacles (i.e., Unavailability of IPs, Unavailability of Support Equipment, and Unavailability of Training Support Areas) suggests that remedial actions other than the addition of time can also be taken to facilitate the ARNG aviators' ability to meet the requirements.

REFERENCE

Cohen, J. (1977). Statistical power analysis for the behavioral sciences (rev. ed.). New York: Academic Press

An Analysis of Differences in Task Accomplishment Related to Mode of Presentation, Live Versus Videotape

Roy Eichhorn
U.S. Army Management Engineering Training Activity

Abstract

Between January 1983 and March 1985 a team from the Army Management Engineering Training Activity (AMETA) conducted an extensive examination of managerial competency among Army civilians. As part of this study, competency assessment inventories were administered to managers at all levels, in all career fields, and in every major command and level of organization in the Army. The results of the study were published as the Army Civilian Executive and Manager Development System (ACE&MDS). Part of the inventory process, involving 1,343 managers, was conducted using two modes of presentation, live and videotape. This paper describes the significant differences in error rates observed with the two modes, proposes an hypothesis for cause, and discusses the implications for standardized presentations of all kinds.

Introduction

Background

The Army Civilian Executive and Manager Development System (ACE&MDS) was designed, prototyped, and tested by the Army Management Engineering Training Activity (AMETA) between January 1983 and March 1985. Briefly, ACE&MDS is a system for assessing an individual manager's job requirements, their importance, and the manager's need for improvement in specific areas. The process is built on a series of managerial competencies developed across all levels of management and from all agencies within the Department of Defense (DOD). The assessment is three-step. First, the position incumbent completes an inventory on his position and himself. Second, the superior completes the inventory on the subordinate. Finally, they get together and reconcile their perceptions of the importance of competency clusters for the incumbent's job, the frequency of performance of the competencies, and whether or not the subordinate needs training in any of the competencies. This last then becomes the manager's development plan. ACE&MDS also includes a system of identifying, categorizing, and matching developmental opportunities to identified needs, either by individual or group.

A major part of the effort in the ACE&MDS design and prototype phases was in the development of a managerial competency assessment inventory. The design phase included an examination of test/retest reliability, and comparisons of inventory versus interview results with managers rated exceptional and those rated fully successful. Once the basic inventory design was established, a low-scale test was organized to collect data from a representative sample of Army managers.

Low-Scale Test

The low-scale test of the inventory was conducted between July and September 1983. To ensure a representative sample of participants, samples were selected from installations in different geographic areas, from every Major Command (MACOM), and from all organizational levels within the command. The stratification factors led to a sample population of 2,400 managers at 42 different installations.

Originally, we intended to have one of the ACE&MDS design team members explain the inventory process, hand out the materials, and then act as a process observer. The process observer role involved collecting data on how long it took to administer the inventory and what kinds of problems people encountered. When the inventory was complete, the team member interviewed participants to elicit the ease of use, their attitudes toward manager development, and related areas. The initial observer team consisted of five people. This was reduced to three before the test began and later, dropped to two. With so many installations to visit, the team considered alternatives to having a team member at each inventory site. The final solution was that we would make a videotape explaining the purpose of the inventory and giving instructions on filling it out. The videotape was shot in one take in an unoccupied conference room, with the presentation done by the project principal investigator. No editing was done. In order to maintain as much control as possible, the team members viewed the tape several times and used the same script when they went out to make live presentations.

Problem

When the inventories came back from the field, records were kept of the inventory spoilage rates and reasons for spoilage. As the inventories were being processed, it became apparent that there were differences in the spoilage rates, installation by installation. The following is an examination of the difference in spoilage rate based on whether the instructions were delivered live or on videotape.

The hypothesis to be tested in this problem are: Ho: the observed difference in spoilage between live instruction and videotape are not significant; H_1: the difference in spoilage rate between live and videotape instruction is significant.

Analysis

In conducting the statistical analysis, data was only included from those installations where the same person who was on the videotape conducted the live presentation. As stated above, the live instructions were given from the same script used for the videotape.

The total sample under consideration was 1,351; 942 live and 409 videotape. The spoilage rates from the two groups were 113 for the former group and 45 from the latter, giving: $\bar{p}$ Live = 7.5 spoiled ballots per installation and $\bar{p}$ Video = 5.6 spoiled ballots per installation.

A t-test conducted on this sample indicated that the Hypothesis of no difference was rejected at the .02 level of significance. We concluded from this that the higher spoilage rate that occurred where live instructions were given was not a chance occurrence.

Possible Causes:

One third of AMETA's mission is teaching, and we have observed that students often appear to pay more attention to videotapes than they do to live instructors presenting the same material. There are time limits to this but it generally holds true for good quality tapes of an hour or less in length. We have hypothesized that the managers participating in the inventory process paid more attention to the TV screen because of the feeling they would only see it once and that they couldn't ask questions. Participants were allowed to ask questions at those sites where there was a team member but this did not improve the spoilage rate above the videotape. At the two installations that had the videotape and had the highest spoilage rates, most of the errors were of only one kind and suggest the intervention of a facilitator. If the spoilage rates for those installations is discounted, Ho can be rejected at a level of significance of .01.

Conclusions/Implications

This analysis has several implications:

1. Videotapes are not only less costly than large numbers of live, standardized presentations but can also be more effective.

2. Based on our analysis, the types of error tend to be fewer with videotape. Therefore, videotaped instructions can be tested, concentrations noted, and the most common types of error reduced. In our case, the most consistent cause for spoilage among the videotape population was improper item ranking. A new videotape could be made with just the explanation of ranking changed. If that problem were corrected, 50% of the spoilage from the videotape population would be eliminated. The pattern for the live presentation is elusive and there is no one area we could correct with consistency.

This analysis suggests that videotape, and related video-disc, or computer-aided instructions, is a viable and possibly a better alternative to sending live specialists into the field to make standardized presentations. In fact, it can be both cheaper and more effective than training trainers. This has considerable potential application for a variety of areas involving new systems or introducing new procedures. In areas outside of manager development, this data suggests that videotape et al. may

not be just a cheap stop-gap for teaching new equipment maintenance and operating procedures, but may in the long run produce better results.

AMETA is continuing its work in non-traditional instruction and competency assessment, and we welcome readers' comments or questions. Address them to: Director, AMETA, ATTN: AMXOM-MT, Rock Island, IL 61299-7040, on phone (309) 782-4041.

Effects of Confinement, Social Isolation, and Diurnal Disruption on Crew Adjustment and Performance on Long Duration Space Missions

Michael D. Matthews, Panel Chair

Department of Behavioral Sciences and Leadership
United States Air Force Academy
Colorado Springs, Colorado 80840

Abstract

A major concern facing planners for future space missions involves the psychological adjustment of crew members under conditions of long duration flights. Effects of prolonged social isolation, confinement, and diurnal disruption on cognitive task performance, social adustment, interpersonal relations, and clinical manifestations are of concern as missions increase in duration from a few days to several months or years. The purpose of the current research project is to investigate a number of areas of human performance and adjustment that may be adversely affected by long duration missions, to assess the impact of these possible decrements upon mission effectiveness, and to explore ways of ameliorating these possible effects. This panel addresses these issues from a multidisciplinary perspective. Presenters and their topics (all of the topics relate to effects of confinement, isolation, and diurnal disruption) are summarized below.

Presenter	Topic
Gary Coulter USAFA	Historical perspective of human considerations in spaceflight
Michael D. Matthews, USAFA	Memory
John Micalizzi USAFA	Attention, perception, and vigilance
Charles Winstead & Greg Smith, USAFA	Cognitive task performance
Len Gardner Consulting Psychologist, Houston, Texas	Social psychological factors in adjustment
Ann Musselman and Patricia Pirrello, USAFA	Stress and its management
Jose Bolton & Jeff Austin, USAFA	Role of physical exercise in mediating stress under conditions of isolation and confinement
Robert Ginnett, Frank Wood, & Thomas McCloy, USAFA	Discussants

Recent Developments in Team Training and Performance Measurement

Chair: Albert S. Glickman, Old Dominion University
Discussant: Bert King, Office of Naval Research

This program has been developed to describe the unique and complex problems/issues associated with team training and peformance measurement and recent developments directed toward addressing the same. Each presentation describes a different aspect of the current approaches to team training.

The first paper serves introduce the issues which differentiate team training from individual training, especially as they impact military performance and readiness. In addition, it will serve as a brief review of the current state of R&D in military team training and the application (or lack thereof) of the results of that R&D in field settings.

Despite the frequent call for further research in team training and performance measurement, there is an extensive body of both published and unpublished literature dealing with team performance. Paper 2 describes a meta-analysis project which is designed to identify gaps in the literature and to integrate and summarize the results of these studies into a usable basis for future recommendations. Initially, over 2300 astracts were identified. Of these, approximately 300 were included in the study domain.

The third paper addresses the practical issues inherent in team training and performance measurement and the current Navy approaches toward these areas of research. As technology increases the demands placed upon Navy teams, the necessity for developing theory-based guidelines for team training becomes more and more critical. This paper describes the current and future development of trainers which include the capabilities for the measurement and enhancement of team performance.

The final paper describes a program of research that aims to identify, clarify, and analyze the interactions that take place among team members as they take place over time. Through this process, recomendations and guidelines will be created for future and current team training. Phase 1 of this research program has resulted in the development of a model of team development which helped to conceptually guide the research program.

Overview of Team Training in the Military

Raymond H. Kirby and Wanda J. Campbell
Old Dominion University

The report of the Defense Science Board Task Force on Training Technology (Alluisi, 1976), calling for a high priority for R&D in crew, group, team, and unit training, crystallized the recognition of the limitations of our research-based knowledge about how teams function and how their training can be accomplished most effectively. A subsequent review by Dyer (1984) reinforced the call for increased R&D on military teams and team training.

For the purposes of this paper and for at least some of the others to follow in the symposium, a team is a distinguishable set of two or more individuals who interact interdependently and adaptively to perform diverse coordinated functions directed toward the achievement of specified, shared and valued objectives. This definition, like that proposed by Dyer (1984), differentiates a team from a small group in that the participants in the latter do not necessarily have assigned roles or functions and coordination of effort, though desirable, is not essential.

Among those issues addressed which have unique applications to teams are the effects of feedback and the effects of structure and workload. Reinforcement contingencies are more complex within the team context than in the individual situation in that reinforcement may be provided to the team as a whole or may be addressed toward individual members as well as the team (Dyer, 1984). An unfortunate consequence of the former situation is that some members may be reinforced for inappropriate responses while others may not be reinforced for correct responses.

As is the case with individuals, high workload conditions are associated with diminished team performance. Increased load impacts on the individual performance of members as well as others in the team who are dependent on the overloaded member (Dyer, 1984). This has resulted in an effort to reduce the coordination required in designing work structures.

In conclusion, it appears that little of the technology that is actually being applied is the result of R&D on how teams function and how they develop proficiency. As this is largely the result of insufficient scientific knowledge, what is needed is a better understanding of how and why teams develop and change, how to measure these changes, and how to intervene to produce the changes we want. Futher, we need to recognize that teams with different tasks, structures, and individual differences among its members may undergo change and respond to interventions differently. Both the task and the opportunity of researchers who confront these issues are considerable.

References

Alluisi, E. A. (1976). Lessons from a study of defense training technology. Journal of Educational Technology Systems, 5(1), 57-76.

Dyer, J. L. (1984). Team research and training: A state-of-the-art review. In Human Factors Review, 285-323.

Impact of Task and Work Structure on Team Performance

Terry Dickinson
Old Dominion University

Eduardo Salas
Naval Training Systems Center

Sharolyn A. Converse
Old Dominion University

Scott Tannenbaum
SUNY at Albany

Recent narrative reviews of the team performance literature (Dyer, 1984; Denson, 1981) have agreed that few prescriptions for enhancing team performance have been generated. Therefore, the objective of this study is to conduct a meta-analysis of team performance research studies in order to aid in the development of techniques with which to investigate and measure team training and performance. Over 2300 abstracts of team performance studies were identified in an extensive search of published and unpublished literature. Each abstract was then examined by three independent raters to determine whether it should be included within the study domain. Approximately 300 studies were admitted to the study domain.

Based on the initial literature review, a meta-model was developed to describe the determinants of team performance and to specify the study domain. A study domain that included research which concentrated on task, work, individual, team and training variables was defined. The meta-model also served as the basis for the development of a code sheet. For each variable of the meta-model (e.g. task type), items and response categories were written to operationalize that variable. Four indepedent raters then coded 30 articles in order to evaluate the availability of data which related to each individual component of the meta-model. The process also served to train coders and to develop a common frame of reference within which to conduct the actual coding of articles. A code book which contained precise definition of potentially vague concepts was also developed to increase interrater reliability.

Currently, research studies are being coded. Study statistics will be calculated to identify correlations between specific team characteristics and measures of team performance. In addition, a number of meta-analyses will be integrated into a meta-analysis of the meta-analyses. This process will serve to determine the magnitude of the causal relationships, refine the meta-model, and identify gaps in the literature. A causal model of team performance which will serve as the basis for future recommendations will be the final product of the project.

References

Denson, R.W. (1981). Team training: Literature review and annotated annotated bibliography. (Report No. AFHRL-TR-80-40). Wright-Patterson Air Force Base. OH.: Logistics and Technical Training Division.

Dyer, J.C. (1984). State-of-the-art review on team training and performance Fort Benning, GA.: Army Research Institute Field Unit.

RESEARCH AND DEVELOPMENT IN NAVY TEAM TRAINING SYSTEMS: TECHNOLOGICAL AND BEHAVIORAL APPLICATIONS

Eduardo Salas and Arthur Blaiwes
Human Factors Division
Naval Training Systems Center, Orlando, FL

The Navy (as well as other DoD agencies) invests major resources and effort to teach members of battle groups how to work together to achieve common objectives. Training of battle group team members generally is enhanced by simulation-based instruction on individual operator tasks, through simulation training for sub-team, single-platform, and finally multiple platform teams. Subsequent to simulator training, team members train as a total battle group using operational equipment at sea, interspersed with additional training in simulation on shore.

In spite of all this apparently systematic training, training specialists criticize team training efforts and programs for offering little direct support for improving team work rather than practice of individual skills in a team environment. Instructors are forced to improvise team instruction where critical team task elements and team interactions are poorly defined. Thus, team skills are trained only incidentally on team trainers. This resulting in reduced training and performance for both individuals and teams.

Clearly, interactions among team members are different from and more complex than interactions between an individual and the equipment. This demands the identification of new and different task elements and performance requirements at the interpersonal level as part of the team training process. This type of team performance diagnosis is being incorporated in two major training systems currently under development at the Naval Training Systems Center: the Surface ASW Training System (Device 14A12) and the Tactical Team Trainer (Device 20A66). These promise significant advances in implementing solutions for team training design problems. Examples of such developments and application will be provided.

The projects outlined in the previous papers comprised significant efforts to improve team training in the military by developing techniques for identifying and teaching tasks that are learned best in team situations as well as for defining and measuring team behaviors via an in situ longitudinal framework essential to the performance of those tasks. These analyses will provide the basis for designing instruction (e.g., practice, cuing, feedback, reinforcement) and instructor aids (e.g., automated performance diagnosis, training guides, behavioral checklists) for teaching team skills. Emphasis is being placed on the development of standardized, objective, and relevant measures of specific team skills to supplement summary outcome measures which say little about acquisition of the component team processes.

The purpose of this paper is to highlight the practical issues inherent in team training as technological and behavioral applications are incorporated into team training systems.

Team Evolution and Maturation (TEAM): An Innovative Approach to Navy Team Training and Performance Measurement

Elizabeth A. Woodard, Ben B. Morgan, Jr.,
R. Craig Montero, & Seth Zimmer
Old Dominion University

An ongoing 3-year project has been initiated to study Team Evaluation and Maturation (TEAM) processes in the operational Navy environment. The purpose of this research is (a) to identify intra- and extra-team variables associated with successful skill acquistition and performance and (b) to conduct longitudinal *in situ* documentation of these processes. The ultimate goal is to determine the optimal design for team training in order to facilitate the acquisition of team skills and to provide valid concepts and operational procedures to enhance operational team performance.

This research is based on the assumption that delineation of the forces and patterns which occur during a team's lifecycle is essential for the effective organization and management of a work group. Moreover, such knowledge will be valuable in determining the specific intervention strategies for use by commanders, managers, planners and trainers to facilitate team development and effective performance. Currently, little rationale exists for choosing one intervention over another. Nor is there a data base capable of guiding decisions at different phases of team maturation.

Based on a review of the group development literature, a preliminary model of the stages of team development has been formulated and is undergoing testing and revision. This model describes several developmental stages through which task-oriented teams may pass. it is believed that the time spent in each stage may vary from team to team. It also illustrates that, there is generally a dual goal of team development: (1) task skill development and maintenance, and (2) team skill development and maintenance. As the two become focused and converge, team performance improves and ultimately is maintained.

The first year of a conceptualized three- part research plan is currently nearing completion. The research has identified the various skills and processes that result in effective team performance, and team training procedures that contribute most to the efficient development of those skills and processes. Data is being collected in order to provide: (a) measures of team-skill levels along several dimensions during different phases of training, and (b) specific guidance for training interventions to correct deficiencies and enhance performance. To make the results most applicable to the Navy, operational teams in training are being used as the test bed of current and future study.

LEADERSHIP PREPAREDNESS IN NEWLY COMMISSIONED NAVAL OFFICERS

Leanne Atwater
Robert Penn
Linda Trumbo

Navy Personnel Research and Development Center[1]
San Diego, California 92152-6800

The degrees to which naval officers were trained to assume leadership responsibilities upon commissioning and after completing Surface Warfare Officer School (SWOS), were assessed. Results indicated that (1) leadership training had a positive effect upon preparedness to become leaders, (2) the Naval Academy generally prepared new officers better than the other two primary commissioning sources, and (3) 2-week Leadership and Management Education and Training (LMET) provided at SWOS added valuable, interpersonal components to the leadership training process.

Introduction

Background

The growing concern for our national security and the concomitant expansion in the size of the Navy point to an increasing need for effective military leadership as expressed by the Secretary of the Navy. While leadership in general is one of the most thoroughly analyzed concepts to be found in the research literature, military leadership has had considerably less systematic attention. Military leadership, specifically leadership training for newly commissioned Navy officers, is the focus of this paper.

Navy personnel have generally viewed themselves first and foremost as leaders. They see the need for strong leaders--those who have the initiative, courage, and knowledge to think and act in situations where military objectives may not be easily recognized--as more critical than ever before. There is concern, however, that many of the events during the past several decades have eroded traditional leadership values. The result has been an overemphasis on management-based, theoretical frameworks designed by social scientists who do not really appreciate leadership in a military context. Articles on this topic, written by the uniformed military are common. Sarkesian (1985), for example, traces the evolution of the corporate management model back to the McNamara years, which focused on cost-effectiveness and econometrics and shifted the focus from leadership in battle to the pursuit of management goals. He feels that the negative impact of this shift was felt in Vietnam. Byron (1985), believes that in peacetime the leadership demanded in combat situations is forgotten and instead, good managers are rewarded. To summarize, many feel that the Navy is losing its military leadership capability, which will in turn affect battle readiness.

An obvious paradox exists in the Navy community in light of all the concerns voiced when it comes to what should be done to develop these ideal leaders. Prior to the inception of this research effort, the authors conducted numerous interviews with senior Navy officers and a common theme emerged. Leadership training is conducted when everything else is finished--it is the last priority--and many, given their way, would throw it out altogether.

In part, this attitude toward leadership training stems from priorities. To teach leadership, something else must go. But this attitude seems to stem also from a lack of agreement as to what military leadership really is, and how (or even if), it should be trained. This definitional problem can be remedied if leadership is viewed as consisting of three dimensions--management, interpersonal skills, and warriorism. While warriorism may or may not be trainable, the majority opinion of leaders represented in a recently published book about military leadership (Taylor & Rosenbach, 1984) is that leadership skills (i.e., managerial and interpersonal skills) can be learned. This study addresses issues relevant to "teaching" these leadership skills to newly commissioned naval officers.

Current Leadership Training in the Navy

The Navy currently provides some training for new officers to help them assume their leadership role. Each of the commissioning sources has a leadership curriculum for naval officers as part of their education and preparation. The specialty schools (e.g., SWOS) continue the leadership training process.

Until recently, the leadership training program at SWOS was a 2-week course designed as part of the LMET program implemented Navy-wide in 1978 in an attempt to standardize the Navy's leadership training. Since 1978 many of the LMET courses have been dropped or shortened. LMET at SWOS is scheduled to be reduced to a 1-week course in 1986. The reasons for this are varied, but reflect to some extent the attitude mentioned earlier that leadership training is low priority training. It also reflects a desire to shorten the overall training pipeline and to get officers into the fleet sooner.

Purpose

If officer training is to be accomplished in the most efficient manner, it is imperative that the training provided be relevant, effective and not redundant. This study set out to address two primary questions. Does training make a difference? That is, are students learning anything they feel will be relevant to their leadership positions? If so, what are they learning and where, and can leadership training be provided more efficiently?

---

[1]The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the Navy Department.

Method

Sample

The sample for this study consisted of 733 newly commissioned junior officers at SWOS. Four hundred and eighty-three students were questioned on their first day of SWOS about leadership training and preparation at their commissioning source. Ninety-eight students were questioned before and after participating in the standard 2-week LMET course and 155 students were questioned before and after participating in a 1-week, condensed version of LMET. Table 1 presents a description of the sample.

Table 1

Description of SWOS Sample

	First Day of SWOS (N = 483)	Pre/Post 2-week LMET (N = 98)	Pre/Post 1-week LMET (N = 144)
Commissioning Source			
USNA	196	3	81
OCS	37	68	13
NROTC	250	27	50
Academic Major			
Science/Engineering	300	41	99
Humanities/Other	183	59	56
Average Age	23	24	23
% choosing surface as 1st choice community	64	55	50
% intending to make Navy a career			
Yes	25	18	17
No	10	20	16
Unsure	65	62	67

Questionnaires

Four versions of a questionnaire were designed to assess leadership training, leadership preparedness, and self-rated leadership abilities, and to collect a number of biographic and demographic characteristics of the students (e.g., academic major, age, sex, commissioning source). Among the four versions the following topics were also measured: managerial style, achieving style as measured by the Manifest Needs Questionnaire, and action vs. state orientation. These topics will not be addressed in this paper.

Questionnaires were administered in classroom settings by SWOS instructors. Administrators were given a set of instructions to read to the students before filling out questionnaires. Students were assured that all informat n was confidential.

Results

Does Leadership Training Really Make a Difference?

An important question in this study had to do with the merit of leadership training--does it really make a difference? Results indicated that leadership training is important to preparedness. A number of different analyses formed the basis of this conclusion. First, the relationship between new students' reports of their training and preparedness were correlated. As expected, the correlations between these two indicators were quite high (.5 to .7). At least in the students' minds they see their training as related to their levels of preparedness.

Second, analyses indicated that students' perceptions of their leadership training and preparation were more related to their feelings of preparation to go to war if necessary than were their self-rated leadership abilities. Preparation to go to war is not only determined by ability; training has an impact.

Third, a number of open-ended comments provided by new students indicated that they felt they needed more leadership training, especially in the interpersonal aspects, in order to best assume their role as division officer.

Fourth, 98 students were questioned before and after the standard 2-week LMET course about their leadership training and preparedness. On the basis of paired t-tests, 14 of the 21 items measuring different aspects of leadership preparedness showed significant improvements as a function of LMET. Many items had differences greater than a standard deviation. Specifics of these differences as well as differences as a function of the 1-week LMET course will be discussed in a later section.

What Leadership Training is Provided?

Training at the Commissioning Sources

The level of leadership training new officers received at the three major commissioning sources was of particular concern in this study. Table 2 presents a list of the leadership training and preparation topics assessed and indicates the areas where officers felt most and least trained and prepared. In general, officers entering SWOS felt they were trained "to some extent" (the midpoint on the scale which ranged from 1, "not at all," to 5, "to a very large extent") in most aspects of leadership. The overall levels of perceived preparation were somewhat higher ($\bar{x}$ = 3.6).

The levels of training differed across commissioning sources in 18 of 21 areas. For the most part, Naval Academy (USNA) graduates felt better trained than those from Naval Reserve Officer Training Corps (NROTC) or Officer Candidate School (OCS), and OCS graduates generally felt least well trained. A noteworthy exception to this was training in how to interact with chiefs in the division. In this area, NROTC graduates felt best trained and Naval Academy graduates felt they had received little training.

The aspects of leadership in which new SWOS students, across commissioning sources, felt they had received the most training and were most prepared were "managing time" and "setting priorities with a heavy workload." Naval Academy graduates, in addition, felt well trained and prepared in terms of managing stress, setting goals and planning work. Graduates of OCS and NROTC, however, felt more prepared in areas of listening effectively and communicating effectively with people than they did in setting goals and planning work.

Areas where new SWOS students felt least prepared were "relieving the division officer" and "performing paperwork requirements." Those from OCS also reported little training in terms of "counseling subordinates," although they felt prepared to some extent to do this. Although not reflected in Table 2, when examining the levels of training and preparation on the more interpersonal aspects of leadership (e.g., knowing how to mo vate subordinates, briefing superiors, counseling subordinates), the levels of training were rather low, and the levels of pre ation were moderate. This was especially true for graduates of NROTC and OCS. It appears that in general, students aving the commissioning sources feel better prepared in terms of managerial type skills (goal-setting and planning) t n in the interpersonal aspects of leadership (counseling and disciplining).

LMET Training at SWOS

The results of the 2-week LMET training improve this state of affairs to some extent (see Table 2). Leadership areas which showed the greatest improvements as a result of this LMET course were "motivating enlisted personnel," "talking comfortably before a large group," "counseling poor performers," "interacting with chiefs in your division," "rewarding and disciplining subordinates," and "relieving the division officer." Also of interest were two items in which perceptions of the levels of training decreased from time 1 to time 2. These areas were "understanding Navy procedures and protocol" and "performing the paperwork requirements as a division officer." It seems likely that LMET served as a realistic preview and made officers aware of the large amount they didn't know in these areas.

The condensed 1-week LMET training at SWOS had different results. Two types of analyses were done to assess the merits of the 1-week course. First, similar to assessing the 2-week course, t-tests were done comparing officers' training and preparation in 21 leadership areas before and after completion of the shortened LMET course. These analyses revealed no improvements in preparedness as a function of the 1-week LMET course. Students did see some improvements in their levels of training. The greatest improvements were in the areas of "taking responsibility for a division," "relieving the division officer," "counseling poor performers," and "rewarding and disciplining subordinates."

A second set of analyses were done to get a picture of the <u>relative</u> pre-post improvements in training and preparation as a function of the two different LMET courses. Analyses of covariance were done on each of the 21 leadership training and preparedness items. Commissioning source and pre-LMET scores were used as two covariates and differences in post-LMET scores were assesed. These analyses revealed no differences in the 1-week vs. 2-week post-LMET training and preparedness scores when the two covariates were taken into account. Essentially, officers leaving 1-week LMET felt as well trained and prepared as officers leaving 2-week LMET. Close examination of the data for these two groups revealed large differences in officers' perceptions of training and preparation at the commissioning sources. It appears that the class members chosen for assessment in the 2-week LMET condition had particularly low perceptions of training at their commissioning sources, primarily OCS and NROTC. This was not the case for two subsequent classes entering SWOS. Essentially, because of the inequities in our one- and two-week LMET samples, a meaningful comparison cannot be made. Students in the 2-week course showed greater improvement as a function of LMET but they had more room for improvement than the one-week class at the outset of LMET.

Self-Ratings of Leadership Ability

While the majority of students' perceptions of their leadership training differed significantly across commissioning sources, self-ratings of leadership ability in eleven areas did not differ. The area in which all students felt most capable was in "doing whatever it takes to get the job done." Aspects of leadership in which students felt least capable were "speaking comfortably in front of a group," and "motivating subordinates to do jobs they don't want to do." In general, officers feel more able to perform managerial duties than to handle the more interpersonal aspects of their leadership role.

The leadership ability ratings of Academy graduates were somewhat surprising. While Academy graduates felt better trained and better prepared than graduates of the other two commissioning sources, they did not rate their leadership abilities any higher. This seems to dispel, to some extent, the stereotype suggested in the interviews that Academy graduates have trouble adjusting in their first tour as division officers due to an over-confidence. It also suggests that abilities are not the only factor relevant to how well prepared officers feel they are. Training does play a role.

Table 2

Leadership Training and Preparation Topics Measured; Extreme Averages By Commissioning Source: Pre-Post Improvement

	New SWOS Students[a]						Greatest Improvement After 2-week LMET
	USNA (N = 196)		OCS (N = 37)		NROTC (N = 250)		(N - 98)
	Trng	Prep	Trng	Prep	Trng	Prep	
a. Making the transition from the Naval Academy, OCS, or NROTC to the operational Navy							
b. Taking responsibility for a division of enlisted personnel							
c. Understanding Navy procedures and protocol			3.5				
d. Relieving the division officer in your first division officer assignment	2.8	2.9			2.7	3.0	X
e. Knowing how to motivate enlisted personnel							X
f. Performing the paperwork requirements as a division officer (PMS, PQS, etc.)	2.5	2.7	2.5	3.0	2.6	2.8	
g. Managing your time and setting priorities when you have a heavy workload	4.4	4.3	3.6	3.9	3.6	3.9	
h. Talking to a large group of people who work for you							
i. Briefing your superior, or the CO about an issue in your division							X
j. Counseling subordinates about personal matters			2.4				
k. Counseling poor performers			2.4	3.3			X
l. Handling alcohol and drug abuse probems among your subordinates						3.3	
m. Resolving conflicts among your crew members					2.7		
n. Listening effectively				3.9			
o. Managing stress (i.e., lack of sleep, disappointing your boss, overwork, conflicts)	4.1	4.2	3.4				
p. Communicating with people effectively				3.9			
q. Demonstrating concern for your subordinates							
r. Setting goals	4.1	4.2				3.7	
s. Planning work	4.1	4.1				3.5	
t. Interacting with Chiefs in your division	2.7	3.1					X
u. Rewarding and disciplining your subordinates							X

[a]Only areas with highest and lowest average levels of training or preparation are presented in this table.

Additional Findings of Interest

Who or What Influenced Leadership Skills?

New SWOS students were asked the extent to which classroom instruction, examples of military leaders, experience leading others, as well as things they learned before they entered USNA, OCS or NROTC influenced the leadership skills they had acquired. Academy and NROTC graduates felt examples of military leaders and experience leading people were most influential. OCS graduates felt they had been most influenced by things they learned before attending OCS. OCS graduates also tended to feel their levels of training had been low in comparison to their levels of preparation. This supports the conclusion that OCS graduates feel their leadership skills are less a function of their officer training than do those who go through NROTC or the Naval Academy. (This is not surprising when the length of these officer training programs is considered (i.e., four months for OCS as opposed to four years for USNA and NROTC).) All three groups felt classroom instruction was least influential, though they still reported it influenced them "to some extent."

Academic Major and Leadership Preparedness

The Secretary of the Navy, John Lehman, and others have suggested that a strict science and engineering curriculum may be too academically narrow to provide new officers with the well-rounded education they need to be good leaders. To address this question, correlations were computed between academic major (science and engineering vs. other) and students' perceptions of their leadership training, preparation and ability. (These correlations were partial correlations controlling for commissioning source.) No significant relationships were found between academic major and leadership training, preparation or ability. Since the measures of leadership are all self-report, they must be treated with some caution, but nonetheless, no relationships emerged.

Discussion

It was encouraging to discover that students felt their training influenced their levels of preparation to become leaders, and that they believed classroom instruction was useful. It was also interesting that academic major had no relationship to students' perceptions of their leadership preparation or leadership abilities. The stereotype that science majors have narrow academic experiences and, therefore, tend to be less sensitive to interpersonal concerns was not supported.

Of particular interest were the training issues addressed. It appears that each of the commissioning sources are preparing new officers to some extent to assume their leadership responsibilities. The bulk of this training seems to impact managerial skills rather than the more interpersonal leadership skills. LMET at SWOS furthers this leadership training and preparation with a positive impact on the interpersonal dimensions. If the Navy's goal is to provide the most efficient training pipeline, it appears that the commissioning sources would do well to concentrate formal leadership training in the managerial skills, leaving the interpersonal skills to LMET. This would be an improvement over providing minimal training in all areas at both schools. Further, while the length and method of the training experiences differ at the various commissioning sources, they would do well to agree on a standard set of leadership issues to be addressed in leadership training and do them as well as possible in the time frame provided.

The comparison of 2-week and 1-week LMET courses was disappointing. It would have been interesting to assess the relative impact of different course lengths if the two SWOS classes had been more comparable. The impact of the shortened LMET course needs further evaluation. While the findings from this study are based on self-report they are suggestive of issues worthy of future pursuit. This work will be followed up with input from SWOS instructors as to students' leadership abilities, further evaluation of the shortened LMET curriculum and optimally, a one-year follow up of individual performance in the fleet.

References

Byron, J. L. (1985). Warriors. Proceedings, June 1985, pp. 64-68.

Sarkesian, S. C. (1985). Leadership and management revisited. The Bureaucrat. Spring, 1985, pp. 20-24.

Taylor, R. L., & Rosenbach, W. E. (1984). Military leadership: In pursuit of excellence. Boulder, Colorado: Westview Press.

PMM and Active Followership

Wm Timothy O'Connell
United States Air Force Academy

ABSTRACT

The Positive Motivation Model (PMM) is a set of guidelines developed by Major Carl Bryand and Captain Michael A. Rosebush of DFBLC that any supervisor or trainer can apply in training a subordinate. The program I propose would teach subordinates to use the PMM principles as a guide for making valuable inputs to the organization. The subordinate must be able to analyze the work environment (including the boss) in terms of the five steps of PMM. In any environment the subordinate should be able to determine what is expected, what skills are needed, how to obtain and use feedback, how to interpret rewards and punishments, and how to pursue self development. The goal of this program would be similar to those pursued by Andre and Ward (1984) to help the subordinate adjust to the supervisor's leadership style and teach the subordinate how to contribute to the unit goals from the bottom up.

INTRODUCTION

BACKGROUND

Rosebush (1984) focused attention on the development of a productive working relationship through a set of guidelines for the superior. The application of these guidelines by upperclass cadets at the Air Force Academy has met with great success during two summer Basic Cadet Training Programs and the present academic year. In surveys conducted after each summer subordinates rated the superiors as the most motivational factor in the training environment. Also, officer instructors have noticed a shift in the upperclass cadet leadership style to a mature style that is appropriate for use after graduation.

Since supervisors continue to be subordinate to higher supervisors, PMM training can be expanded to include the subordinate. In a dual role the supervisors/subordinates need to know how to contribute to the organization both up and down the chain of command. Whether the superior is using the PMM style is not relevant since an active subordinate can effectively exercise self motivation and positively influence others as an active follower.

A common saying is that good followers make good leaders. A corollary is that former leaders make good followers. They have seen the impact of strong, active followership. This expansion of the PMM training program to subordinates should develop these positive, active followership skills.

THE PRINCIPLES OF PMM

To understand the principles of PMM as they apply to the subordinate you must first understand the five parts of PMM as they are taught to the supervisor. First the supervisor must establish expectations. This includes introducing oneself by stating one's position, background, and values. It also includes setting a positive atmosphere in which the superior will help but not carry

the subordinate, the superior appreciates the subordinate's strengths and feedback, and the superior expects commitment to the program. When setting expectations the superior also explains the rationale for the task and provides an overview of the future.

The second part of PMM is providing the subordinate with the skills necessary to accomplish the task. The task often dictates the training method employed, but some universal principles do apply. For example, the superior should determine the learning objectives, demonstrate the skill by example, and supervise the subordinate's rehearsal of the skill.

The third part of PMM is providing feedback. The rules for feedback are summarized by the acronym "INPUT +". It stands for immediate, no labeling, proper person (addressed specifically to a subordinate), uniquely specific (give specific details about the behavior which is the topic of the feedback), talk about the behavior (and not the person) and end with a positive, resupportive statement.

The fourth part of PMM, consequences, deals with rewards and punishments. Superiors are taught to make consequences immediate, consistent, a resultant of the behavior, progressively building, and meaningful from the subordinate's viewpoint. They must also learn how to provide them. The superior is taught that the purpose of punishments is to direct behavior, to teach and help the subordinate to understand such things as priorities, to back up the supervisor's expectations, and to reaffirm the supervisor's commitment to standards.

The last part of PMM is growth. The goal is to increase the subordinate's ability to accomplish a task with less supervision by increasing his or her self esteem, by establishing performance goals, and by establishing homework assignments.

With this overview as a foundation for understanding PMM let's look at each of these five parts from the subordinate's viewpoint. Our goal changes now from maximizing the supervisor's leadership skills to maximizing the subordinate's positive, active followership skills.

EXPECTATIONS

Understanding what is expected of the subordinate is as much the subordinate's responsibility as it is the supervisor's. The subordinate should start by asking a series of questions which parallel the PMM outline for setting expectations.

From the minute the superior first introduces him or herself the subordinate needs to start asking these questions. What is my boss's background? What does he or she not know that I do? How can I help? How should my boss's values effect my priorities? When do I make a decision? What should I bring to the boss for inputs before making decisions? Does the boss know the strengths I bring to this job? What route should I use to provide feedback? How important is my current task relative to others? What factors will change in the future?

The objective of this phase of training is to teach the the subordinate to

become actively involved in the work environment. The subordinate must aggressively seek to understand the boss, the importance of the task, the course of the future, and his or her expected level of commitment.

FEEDBACK AND CONSEQUENCES (and SKILLS)

Unlike the approach used in teaching PMM to supervisors, we do not divide the discussion of feedback and consequence. We consider them together in two different situations: dealing with the feedback and consequences directed to the subordinate and giving the organization, specifically the boss, feedback and consequences.

When receiving feedback you, as the subordinate, learn how to evaluate its contents against the PMM acronym INPUT +. You must get the feedback immediately. The superior should not label but if you receive one you must fight the emotional reaction with self reassurance. Determine which person the boss is talking to. If it's not you, you should still learn from the situation since the boss is trying to set an example. Get the unique specifics concerning your behavior. Determine the exact causes for the good or bad outcome. Differentiate when the boss is talking to a person he may or may not like and when he is talking about your behavior. And finally, use your boss's positive encouragement to reinforce your motivation.

For cases of feedback and consequences for negative behavior you are taught to take an active role in his or her situation. Be immediate in dealing with it. Analyze what you did without alibiing, blaming others, or self labeling. Determine specifically what caused the problem and develop a game plan for improvement. Ask for more training (skills) if needed. Use the boss's positive reassurance as a springboard for future improvement. As Anthony (1983) points out, be a problem solver, not a creator, by bring the boss in on the game plan.

When the subordinate has done well he or she should get involved in the reward process. Andre and Ward (1984) outlines a way of doing this. The subordinate should set goals and determine the milestones needed to obtain those goals. Later, when the boss determines that a reward is due, the subordinate may request an appropriate reward which would meet his or her milestone. To be an appropriate reward it must be consistent with the boss's award policy and equivalent to the behavior which produced it. Furthermore, the subordinate must be infrequent in proposing a reward. The act of rewarding must still belong to the boss.

If the consequences are due to negative behavior, the subordinate must learn to approach it as a learning process. He or she has to determine what behavior produced the consequence, to formulate a game plan for improvement, to reevaluate the boss's expectations, and to reassure the boss, through actions, that the behavior was temporary.

Giving feedback to the boss follows a different approach. Feedback from the subordinate must be provided to ensure that the boss can make informed decision and to ensure that the organization can function properly. Anthony (1983) cautions the subordinates against apple-polishing, but points out that sincere praise from a trusted subordinate can be a meaningful reward for a boss. In the case of negative feedback the key factor is to build a favorable

reputation. Crockett (1981) breaks this process into a number of behaviors that build trust in the relationship. The subordinate is not only taught these concepts but also the need to start developing a trusting relationship with the boss well before any feedback from the subordinate is needed. Once this is accomplished the subordinate can use a variety of techniques which communicate a need for change.

Consequences for the boss are limited. Negative consequences are not taught and the subordinates are cautioned against using any they may discover. However, subordinates are taught a wide range of rewards they can use to strengthen the boss's leadership of the unit and promote unit success. The subordinate is shown how to reward the boss through extra projects, overtime, etc.

GROWTH

This is the most important section of the training. A common theme has prevailed throughout the training: self improvement and unit goals can be mutually supportive. This theme is refined with specific steps the subordinate can follow to assimilate to a new work environment, to improve the subordinate's strengths and use them for the unit's benefit, to alleviate weaknesses, to self evaluate past performance, to correct or reinforce behavior, and to establish goals. This is taught mostly via personal case studies. The subordinates analyze various situations they have been faced with. They are asked to talk about themselves in the third person through the boss's perspective. Then they must make a public statement of the growth goals. As research by Anthony (1983) indicates, goals set publicly have a higher rate of accomplishment than those set privately.

GOAL

The goal of this entire program is to ensure the subordinates can grow within the organization with mutual benefit to both themself and the organization.

REFERENCES

Andre, R. and Ward, P. (1984). The :59 Second Employee. Boston, Massachusetts: Houghton Miffin

Anthony, W. (1983). Managing Your Boss. New York, New York: American Management Associations

Crockett, W. (1981). Dynamic subordinate. Training and Development Journal, May 1981.

Rosebush, M. (1985). Applying the Positive Motivation Model (PMM). USAF Academy, Colorado: GPO

THE POSITIVE MOTIVATION MODEL: ITS USE AT THE AIR FORCE ACADEMY

CAPTAIN LYNN M. SCOTT
LT COL WILLIAM H. CLOVER

DIRECTORATE OF LEADERSHIP PROGRAMS AND STUDIES
DEPARTMENT OF BEHAVIORAL SCIENCES AND LEADERSHIP, USAF ACADEMY COLORADO

ABSTRACT

This paper discusses the implentation and use of the Air Force Academy's new method of leadership training, the Positive Motivation Model during the 1985 Basic Cadet Training Program. The model's effectiveness was measured through a survey of subordinates perceptions of their supervisors uses of the principles. Comparisons are made among basic cadets, upperclass cadre and across the two Basic Cadet Training phases. Future enhancements are also discussed.

INTRODUCTION

During the past two years, the Air Force Academy has implemented a new leadership training program called the Positive Motivation Model or PMM (Rosebush and Bryant, 1984; Rosebush, 1985). PMM focuses on five principles of leadership: the leader's communication of expectations to subordinates; the leader teaching necessary skills to his or her subordinates; providing feedback about performance to subordinates; establishing a contingency schedule of rewards for desired behaviors and consequences for undersirable behaviors; and providing an overall climate and sense of growth for subordinates.

Cadets in the classes of 1985 and 1986 were given extensive instruction on the PMM philosphy and how to incorporate it into their leadership styles. The first test of PMM's utility came in the summer of 1984 during Basic Cadet Training (BCT) for the class of 1988. Basic Cadet Training is the freshman class's first exposure to the military rigors of the Air Force Academy. The program is administered into two three week phases aptly called 1st BCT and 2nd BCT. The first phase focuses on fundamental military indoctrination (e.g. marching, rifle manual, room and uniform inspections, intramural athletics and Air Force Heritage). The second phase is largely a field encampment where cadets live in tents, negotiate numerous obstacle courses, learn land navigation and land battle tactics as well as marksmanship and unarmed combat. All basic cadets are organized into 10 squadrons which are run by upperclass cadre. According to a survey administered by the Academy's Office of Institutional Research to the class of 1984 during BCT this first attempt to use PMM was successful. The cadre were identified as major factors in motivating basic cadets to continue through the BCT program. In the past, cadre had been identified as contributing to the desire to leave the Academy (McFarland, 1985).

During the next 12 months, Rosebush (1985) developed the Positive Motivation Model Inventory, PMMI, to provide a more rigorous assessment of the overall level of positive motivation among subordinates and to operationally define and measure the model's five concepts. The PMMI consists of items that were heuristically organized into six scales; one for each of the PMM concepts and one for the construct positive motivation.

The purpose of this paper is to discuss the administrations of the PMMI to the basic cadet class of 1989 and to cadre members in the summer of 1985 to assess the use of PMM at the superior to subordinate level. Rating comparisons will be made between each of the two phases of BCT and between the cadre's use of PMM among themselves and with the basic cadets.

METHOD

SUBJECTS

The subjects consisted of 1305 basic cadets from the class of 1989 and 525 upperclass cadre from the classes 1986 and 1987 who were responsible for conducting the BCT program.

APPARATUS

The subjects were surveyed with the Positive Motivation Model Inventory (PMMI consisting of 68 questions and a seven-point Likert Rating Scale). Data were gathered on Westinghouse Learning Center optically scannable answer sheets.

PROCEDURE

The PMMI was administered at the end of each BCT phase in mass to all basic cadets and individually the cadre members. The survey instructions explained that the questions in the PMMI were designed to measure the subordinates' perceptions of how well their supervisor is applying the principles of supervision espoused in the Positive Motivation Model. The instructions continued, saying that the source of responses would remain confidential and that their candidness would be essential in helping their supervisor learn about his or her strengths and weaknesses. All basic cadets were instructed to rate their element leader while the cadre rated their immediate cadet supervisors. Analyses were conducted using the Statistical Package for the Social Sciences (SPSS) programs on the Burroughs 6900 mainframe.

ANALYSES

The analyses addressed three questions about the PMMI and the level and stability of the PMM scale scores. They were: (1) assessing the internal consistency of the heuristically developed PMM scales; (2) determining whether there was a consistent use of PMM concepts between element leaders and basic cadets during each BCT phase; and (3) determining whether there was a consistent use of PMM concepts between cadre supervisors and cadre subordinates during each BCT phase.

RESULTS

Internal consistency was calculated using Cronbach's Alpha. Items detrimental to internal consistency were eleminated from each scale. Resultant alpha values ranged from .93 for the scale called Positive Motivation to .77 for the scale called Skills. Figure 1 lists the items that comprise each scale and their respective alpha values.

One-way analysis of variance comparing basic cadets' ratings on all six scales across 10 squadrons during 1st BCT did not identify any significant mean differences. However significant differences were found on all scales during 2nd BCT. For the cadre, the analysis of variance revealed significant differences in all but the scale called Feedback for 1st BCT but significant differences were only found for two scales, Skills and Consequences with 2nd BCT cadre data.

T-Tests were also performed to compare mean scale scores across BCT phases for basic cadets and cadre. The basic cadet comparisons revealed significant differences on all scales but Feedback. Cadre comparisons did not reveal significant differences on any scale. These results are summarized in Table 1.

DISCUSSION

First of all, despite any reported differences between means, the high ratings indicate that the Positive Motivation Model was being used well by the BCT cadre. The lack of significant differences on any of the PMM scales for basic cadet during 1st BCT may suggest that the total novelty of 1st BCT and cadre leadership styles produces no differences in the perception of how PMM is used. That is, since basic cadets have never experienced BCT before there should be a uniform communication and understanding of expectations, the skills that are needed, and the sense of growth and positive motivation. In turn, the BCT system, which evokes predictable behaviors (both desirable and undesirable) from basic cadets is probably responsible for similar ratings on the feedback and consequences scales. The first BCT phase however acts as both an acclimatizing period and a benchmark for comparing for 2nd BCT activities. This may explain why there are significant differences on all scales during this second phase. In some squadrons basic cadets may have adjusted to the program and have a clearer understanding of expectations, and have acquired more skills than cadets in other squadrons. Differences in feedback, consequences, growth and positive motivation may be caused by negative feelings about the new element leader in 2nd BCT when compared to previous element leader from 1st BCT. Overall this 1st BCT reference comparison may explain why there are significant differences for all but one scale between BCT phases for basic cadets. This perspective however does not adequately explain the pattern of significant differences for the cadre. The cadre members are experiencing BCT from an entirely different perspective. They are largely acquained with the duties they'll perform in either 1st BCT or 2nd BCT. There are differences by squadron for each BCT phase and this could be attributed to different leadership styles of the squadron commander or to each squadron defining cadre roles differently.

Following the analysis, cadre members who had five or more subordinates were sent personalized and confidential feedback about their performance on each PMM scale. The feedback package included: (1) a cover letter commending the cadets on their use of PMM which also explained that only the cadet would have access to his/her scores and provided points of contact within the Department of Behavioral Sciences and Leadership for questions or comments; and (2) a separate sheet containing the cadet's PMM scores on each scale along with cadre averages on each scale for comparison.

The entire PMM program, from instruction to performance measurement and feedback is to the authors' knowledge the first of its kind to be used in a non assessment center or service academy environment. Future activities include periodically surveying the entire cadet wing, providing more comprehensive feedback to cadets and refining the PMMI. The Positive Motivation Model has clearly opened the door for systematic and testable forms of leadership training and established the bridge between leadership theory as taught by the Department of Behavioral Sciences and Leadership and leadership applications and experiences provided by the Commandant of Cadets.

POSITIVE MOTIVATION MODEL: EXPECTATIONS (EXPECT) α = .82
EXPECT = (#97+#103+#111+#116)/4

I have a clear understanding of my primary rater's values. (# 97)

My primary rater genuinely desires to help me do well at my job. (# 103)

My primary rater lets me know the degree of importance behind the tasks he/she provides me. (# 111)

I have no doubt about what my primary rater expects of me. (# 116)

POSITIVE MOTIVATION MODEL: SKILLS (SKILLS) α = .77
SKILLS = (#96+#102+#110)/3

My primary rater presents new information to me in a manner I can easily understand. (# 96)

My primary rater leads by example. (# 102)

I believe my primary rater does an excellent job of making me skilled at the tasks he/she expects me to accomplish. (# 110)

POSITIVE MOTIVATION MODEL: CONSEQUENCES (CONSEQ) α = .83
CONSEQ = (#101+#109+#115+#118)/4

Both rewards and discipline are immediately given to me by my primary rater. (# 101)

When my primary rater disciplines me, I believe he/she is genuinely trying to help me. (# 109)

My primary rater rewards me for showing improvement even though my performance is not yet perfect. (# 115)

My primary rater is consistent with how he/she gives me rewards and discipline. (# 118)

POSITIVE MOTIVATION MODEL: FEEDBACK (FDBCK) α = .82
FDBCK = (#100+#107+#114+#117)/4

When I do something right, my primary rater will publicly recognize it. (# 100)

My primary rater provides me both negative and positive feedback immediately. (# 107)

The feedback I get from my primary rater is so specific I know exactly what he/she likes or dislikes about my performance. (# 114)

My primary rater always provides me with enough feedback to know whether I've done a task correctly or not. (# 117)

POSITIVE MOTIVATION MODEL: GROWTH (GROWTH) α = .84
GROWTH = (#99+#106+#113)/3

I genuinely believe my primary rater supports me. (# 99)

My primary rater has helped make me feel very confident in myself. (# 106)

My primary rater helps me feel successful. (# 113)

POSITIVE MOTIVATION MODEL: POSITIVE MOTIVATION (POSMO) α = .93
POSMO = (#98+#105+#112)/3

I feel a tremendous amount of respect for my primary rater. (# 98)

I would gladly volunteer to work for my primary rater again. (# 105)

If I ever get my primary rater's job, I would want to treat my subordinates the same way he/she treated me. (# 112)

Figure 1. PMM Scale Items And Their Respective Cronbach Alpha Values

Table I

Overall PMM Averages For Basic Cadets And Cadre

	EXPECTATIONS	SKILLS	FEEDBACK	CONSEQUENCES	GROWTH	POSITIVE MOTIVATION
Basic Cadets, 1st BCT	5.40	5.37	5.29	5.5[illegible]	5.51	5.65
Basic Cadets, 2nd BCT	5.29	5.18	5.23	5.[illegible]	5.41	5.31
Cadre, 1st BCT	5.42	5.22	4.98	4.96	5.46	5.33
Cadre, 2nd BCT	5.27	5.08	4.78	4.90	5.31	5.22

Note. N= 1305 Basic Cadets and 525 Upperclass Cadre

REFERENCES

McFarland, B.P., Personal Communication, August, 1985.

Nie, N. H., Hull, C.H., SPSS update 7-9: new procedures and facilities for releases 7-9. New York: McGraw-Hill, 1981.

Rosebush, M. A. and Bryant, C., The positive motivation model (PMM) Air Force Air Force Academy Cadet Counseling Center Presentation, April 1984.

Rosebush, M. A., Applying the Positive Motivation Model (PMM) United States Air Force Academy, Department of Behavioral Sciences and Leadership, Cadet Counseling Center publication, February 1985.

Role Perceptions and Role Adaptations in Hierarchial Dyads

V. Ann Sweney
Wichita State University

Abstract

The Role Reaction Theory hypothesized that individuals assume a role based upon perceptions and pressures inherent in a dyadic relationship. The four instruments utilized each identifies either five superordinate roles or five subordinate roles. Each subject indicated the roles they 'played' in response to their dyadic partner as well as their perception of what roles that same partner was utilizing. The pilot study using 55 dyads composed of Titan and Minuteman Missle and SAC Tanker crews produced promising results. Instrument refinement based upon these results increased the reliabilities to the .74 to .96 range for the twenty scales. The current results were based on 244 dyads drawn from a wide variety of civilian work settings. They gave strong support to the relationships proposed by the model. Twenty-five of the thirty hypothesized relationships between preceptions and role pressures within individuals and across dyads were supported to the .001 level.

Introduction

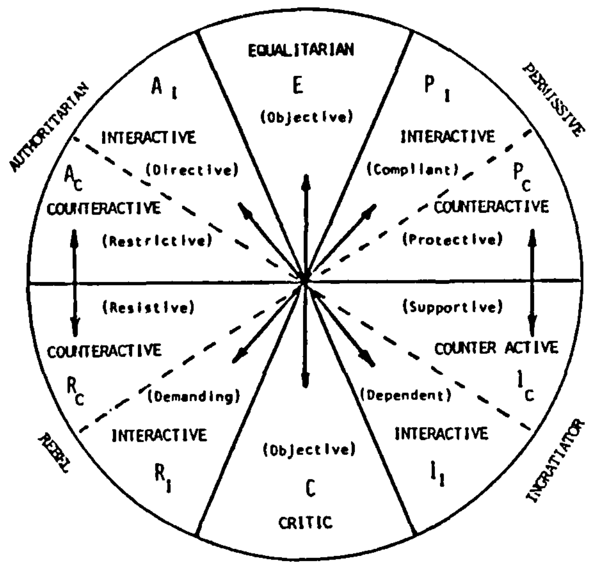
There has been a renewed interest in the concepts and constructs of leadership as organizations have been faced with the need to be more effective and efficient. The original work done by Lewin, Lippett, and White (1939) still provides good basic theory due to the focus on functional roles. Likert (1961), Blake & Mouton (1967), and Fiedler (1967) all provided models against which results could be compared and inferences drawn for success or lack of success. One handicap of these models was the concentrated focus on just the leader without concern about the impact of the subordinate. Leary (1957), Mechanic (1964), and Zaleznik (1958) worked in the area of subordinates but did not incorporate superordinate considerations.

In 1970 Sweney presented the Response to Power Model as the foundation system to analysis superordinate/subordinate relationships. As the title suggests, the underlying theory was the dyadic interaction between individuals in response to the utilization of power and obligation. The model provided three superordinate and three subordinate roles. Of the six roles, four were perceived to be manipulative, Authoritarian and Permissive for the superordinate with Ingratiator and Rebel for the subordinate. The other two roles, the Equalitarian superordinate and the Critic subordinate were designed to be objective-fact oriented roles. Five instruments were developed to operationalize the constructs tapping the individual's value system, the respondent's reaction to specific role pressures, and finally his/her perceptions of the roles played by either their superordinate or subordinate. The ensuing years of research on the effectiveness of the model provide sufficient levels of significance to indicate that the initial concepts were sound.

The RPM model did not identify motivational components involved in role behavior so it did not sufficiently differentiate the kind of relationships that might evolve among the manipulative roles. Sweney (1972) and Sweney (1980) suggested the Role Reaction Model as a possible successor. The new model assumes that all four manipulative roles are differentiated into two motivational components. Reinforcement theory suggests a long-term symbiotic relationships that is "interactive". Dissonance theory measures short-term efforts to change the other person and thus is "counteractive". This has resulted in a model that provides five superordinate roles and five subordinate roles clearly identifying the motivational aspect of relationships. Of the five instruments developed, the four that measure perception and adaptation to the pressures of the role partner were utilized in this study.

Role Reaction Model

Role theory provides the under-pinning for interaction between individuals. Each role seems to have basic accepted norms for behavior in that role and expectations of the behaviors of the role partner. When the roles are reciprocal each individual complements the other and the long-term relationship is interactive. In the Role Reaction Model there are three such interactive dyadic relationships. The Authoritarian Interactive directs, providing structure and organization for the dependent Ingratiator Interactive who feels a need for certainty. This form of dyadic relationship is traditional for most organizations. The Permissive Interactive is compliant to the demands of the Rebel Interactive. This superordinate perceives that the earning of love and cooperations comes from appeasing disgruntled subordinates who never really wants to be satisfied. This Avant-garde relationship results in the Rebel Interactive having the power with none of the obligation since the superordinate assumes the responsibility for making the subordinate happy.



The third reciprocal relationship is between the Equalitarian superordiant and Critic subordinate. This one is different since neither party is focused upon the personal ego needs of the other. Their relationship is based upon the objective handling of facts and situations necessary to accomplish the task. Power and obligation are handled in a rational manner to maximize productivity and individual contributions to the effort.

The last two dyadic role relationships are indicators of conflict or active avoidance of conflict over power. The antagonistic relationship is found between the Authoritarian Counteractive and Rebel Counteractive both of whom want more power. The mutual support relationship between the Permissive Counteractive and Ingratiator Counteractive is manifested by neither party wanting power since it is seen as having detrimental effect on the relationship. In both cases, performance suffers since all four roles have high ego

involvement.

The initial research on the model examined 55 dyads from Titan and Minuteman Missle and SAC Tanker crews. The model was confirmed and a large number of the hypothisized relationships were validated. Since the instruments utilized were in the developmental stage, it was unclear if the weaker statistical results were caused by model conceptualization or the measurement of the constructs. The ensuing years have been spent in instrument refinement that increased the scales reliabilities into the .74 to .96 range. Now that it is perceived that the instruments are sufficiently reliable, the purpose of this study is the validation of the model.

Study

The 244 dyads analyzed in this study were drawn from a wide variety of work settings. The age range for subordinates was 17 to 64 and for the superordinates 18 to 63 with an average of 32 and 40 respectfully. The organizational level ranged from those with no subordinates (0) to a high of level 9. The length of tenure for the relationship measured ranged from less than a year to 54 years with an average of 3 years.

Each respondent of the dyad completed two different instruments with the instructions of keeping the role partner of that dyad as the target for responding. Subordinates answered the Superior/Subordinate Reaction Test (Subordinate Form) to indicate which role they adopted due to the pressures or actions of their superordinate. They also completed the Superiors Purpose Rating measuring the roles they percieved their superordinate to be portraying. The Superordinate filled out the superordinate form of the Superior/Subordinate Reaction Test and the Subordinate's Motivational Rating to supply the same information, adopted role and perception of the subordinate.

The results of the Pearson Correlation are found in the tables below.

Table 1: Adaptation Role Measured by SSRT Scales
Correlated with Perception of Partner's Role Measured by SPR or SMR
N=244

Subject	Role Adopted	Test	Perceived Partner role	Test	r
Sub.	Rebel Counter.	SSRT	Author. Counter	SPR	.406**
Sub.	Rebel Inter.	SSRT	Permissive. Inter.	SPR	.258**
Sub.	Critic	SSRT	Equalitarian	SPR	.578**
Sub.	Ingratiator Inter.	SSRT	Author. Inter.	SPR	.090
Sub.	Ingratiator Counter.	SSRT	Permissive. Counter.	SPR	.428**
Sup.	Author. Counter.	SSRT	Rebel Counter.	SMR	.381**
Sup.	Permissive Inter.	SSRT	Rebel Inter.	SMR	.348**
Sup.	Equalitarian	SSRT	Critic	SMR	.517**
Sup.	Author. Inter.	SSRT	Ingratiator Inter.	SMR	.332**
Sup.	Permissive Counter.	SSRT	Ingratiator Counter.	SMR	.400**

* = .01 significant level ** = .001 significant level

Table 1 illustrates that each respondent role adoption corresponded to the hypothesized pattern while perceiving the role of their partner with one exception. The subordinate Ingratiator Interactive positively correlated with the superordinate Authoritarian Interactive, but not significantly.

Table 2: Accuracy in Perception of Dyadic Partner's Role
N=244

Subject	Perceived Partner's Role	Test	Adopted Partner's Role	Test	r
Sub.	Author. Counter.	SPR	Author. Counter.	SSRT	.369**
Sub.	Author. Inter.	SPR	Author. Inter.	SSRT	.195*
Sub.	Equalitarian	SPR	Equalitarian	SSRT	.404**
Sub.	Permissive Inter.	SPR	Permissive Inter.	SSRT	.305**
Sub.	Permissive Counter.	SPR	Permissive Counter.	SSRT	.191*
Sup.	Rebel Counter.	SMR	Rebel Counter.	SSRT	.472**
Sup.	Rebel Inter.	SMR	Rebel Inter.	SSRT	.320**
Sup.	Critic	SMR	Critic	SSRT	.516**
Sup.	Ingratiator Inter.	SMR	Ingratiator Inter.	SSRT	.324**
Sup.	Ingratiator Counter.	SMR	Ingratiator Counter.	SSRT	.351**

* = .01 significant level ** = .001 significant level

The next question pertaining to the measurement of the model was the level of accuracy with which the dyadic individuals were able to perceive the roles adopted by their partners. By comparing the perception instruments to the SSRT instruments, it was ascertained that the level of accuracy was extraordinary. All nine of the ten correlations were significant to the .001 level while the tenth was significant to the .01 level. (Table 2).

Table 3 answers the question of whether the respondents adaptation role corresponded to the model. This would require them to clearly identify which role behaviors would result from certain pressures applied by a role partner. Once more the Ingratiator Interactive and Authoritarian Interactive failed to result in a significant relationship.

Table 3: Net Role Adaptation as Measured by the SSRT for Each Partner
N = 244

Subordinate's Adopted role	Superordinate's Adopted role	r
Rebel Counteractive	Authoritarian Counteractive	.281**
Rebel Interactive	Permissive Interactive	.155*
Critic	Equalitarian	.266**
Intratiator Interactive	Authoritarian Interactive	.058
Ingratiator Counteractive	Permissive Counteractive	.221**

* = .01 significant level ** = .001 significant level

Table 4: Complementation of Perception Across Dyadic Partners
N = 244

Subs Perception of Superordinate	Test	Super. Perception of Subordinate	Test	r
Authoritarian Counter.	SPR	Rebel Counteractive	SMR	.331**
Authoritarian Inter.	SPR	Ingratiator Interactive	SMR	.076
Equalitarian	SPR	Critic	SMR	.453**
Permissive Interactive	SPR	Rebel Interactive	SMR	.232**
Permissive Counteractive	SPR	Ingratiator Counter.	SMR	.276**

* = .01 significant level ** = .001 significant level

The complementation of perceptions across the dyads, as illustrated in Table 4, once more indicates the non-significant relationship between Authoritarian Interactive and Ingratiator Interactive. In an attempt to understand the dynamics of this dyadic pattern and why the model was not being confirmed, other correlational patterns were investigated. The subordinate who adopted the Ingratiator Interactive role is negatively related to the Permissive Interactive role and positively related to the tenure of the superordinate. This may indicate that being in need of certainty and direction led to the perception that more experienced superordinates, who were not Permissive, would provide the need for structure regardless of specific role.

References

Blake, R.R. & Mouton, J.S., The Managerial Grid, Gulf Publishing Co., Houston, 1964

Fiedler, F.E., A Theory of Leadership Effectiveness, McGraw-Hill, N.Y., 1967

Leary, T., Interpersonal Diagnosis of Personality, Ronald Press, NY. 1957

Lewin, K., Lippett, T., & White, R., "Patterns of agressive behavior in experimentally created social climates", Journal of Social Psychology, 1939, 10, 271-299.

Likert, R., New Patterns of Management, McGraw-Hill, NY, 1961

Mechanic, D., "Sources of power of lower participants in complex organizations', Adminstrative Science Quarterly, 1964, 7, 364-394

Sweney, A.B., "Organizational power roles", Professional Management Bulletin, 1970, 10, 5-13

Sweney, A.B., A Study: Role Reaction Model, AFOSR #2001 Research Grant, 1972

Sweney, V.A. & Sweney, A.B., "Validating the Role Reaction Model for Predicting superior-subordinate relationships", Proceedings:Psychology in the Department of Defense, Seventh Annual Symposium, Colorado Springs, USAF Academy, 1980

Zaleznik, A., "The dynamics of subordinancy", Harvard Business Review, 1958, 43, 18

Differing Attributions of Leadership by Leaders and Their Superiors

Jerry P. Haenisch, Captain, USAF
90th Civil Engineering Squadron, F.E. Warren AFB, WY

Benjamin L. Dilla, Captain, USAF
Air Force Institute of Technology, Wright-Patterson AFB, OH

This study attempted to define leadership by U.S. Air Force Base Civil Engineers (BCEs) in terms of specific BCE behaviors. BCEs and their superiors (wing and base commanders), totalling 160, rated a variety of behaviors as demonstrating poor leadership, good leadership, or having no relation to leadership. Results showed some important differences between the responses of the leaders and their superiors. The superior officers tended to emphasize enforcement of standards, initiation of communication, and support of organizational rules, policies and the chain of command. The BCEs, on the other hand, stressed the importance of people oriented actions within the unit. Behaviors such as recognition of individuals, fairness in personnel actions, and representing and protecting the work force were emphasized. Knowledge of such differences in evaluation of leadership ability could help clarify expectations and stimulate communication between leaders and their superiors.

Leadership is a topic of almost universal interest. It has been especially stressed in the hierarchical military environment. Its relevance to the field of Air Force (AF) civil engineering in particular was highlighted in a 1983 study by McKnight and Parker. These authors found that AF wing and base commanders rated leadership as one of the most important determining factors in the effectiveness of base Civil Engineering (CE) squadrons.

Yet the concept of leadership has not been universally defined. Definitions abound, theories proliferate, and a myriad of research has yet to lay the controversy to rest. Years of study have led many to question whether leadership even exists.

These problems of definition and understanding led recently to an attributional approach to leadership--the idea that leadership is in the eye of the beholder. Pfeffer (1977) stated: "Whether or not leader behavior actually influences performance or effectiveness, it is important because people believe it does" (p. 110). Rush, Thomas, and Lord (1977) found evidence for the influence of personal leadership theories on the responses to leadership questionnaires.

The attributional approach to leadership was perhaps developed best and stated most strongly by Calder (1977). He wrote: "Leadership exists only as a perception. Leadership is not a viable scientific construct" (p. 202). Despite this strong negative verdict on the concept itself, Calder did not call for the end of leadership research, just a change in focus which would incorporate the layman's perspective in evaluating leadership. "If it does nothing more than call attention to the need for understanding the everyday, nonscientific meaning of leadership for specific groups of actors, attribution theory represents an advance for both leadership research and training" (p. 202). This emphasis on perceptions quickly found a home among those engaged in the scientific study of leadership, as reflected by its emphasis in the landmark review of the leadership literature by House and Baetz (1979).

The McKnight and Parker (1983) study indicated a possible link between the wing and base commanders' perception of leadership by the Base Civil Engineer (BCE) and overall CE effectiveness. Informal conversations with top CE officers confirmed that commanders' perceptions of leadership by the BCE are a critical measure. How these commanders define leadership requires clarification before further analysis can be made concerning the leadership-effectiveness link. The present research was an attempt to establish an operational definition of BCE leadership from the perspective of wing and base commanders. Perceptions of leadership behavior by the BCEs as incumbent leaders were also gathered to provide a basis for comparison.

Thus, the present research relies on the attributional approach as its foundation and reflects an attempt to define leadership for a specific group of leaders from the perspective of superiors and incumbent leaders. This approach stands in contrast to the typical reliance on subordinate evaluation of leader behavior. The research was completed as a master's thesis project by the first author, under the advisement of the second author, at the Air Force Institute of Technology (Haenisch, 1984).

Method

Sample

Surveys were sent to wing commanders, base commanders, and BCEs at 86 AF bases in the continental U.S. (CONUS); this represented the entire population of interest.

Procedure

A fairly short survey (7 pages, 57 items) was mailed to each officer at their duty addresses. The survey was accompanied by a cover letter explaining the nature of the research in general terms and carrying an endorsement by the Dean of AFIT's School of Systems and Logistics (a USAF O-6).

Measures

The survey was composed of three demographic items (major air command, base size, and duty title of the respondent), a section for rating 45 possible BCE leadership behaviors, a section for rating nine effectiveness criteria, and spaces for comments regarding BCE actions most damaging to leadership and for other comments about BCE leadership and its measurement.

The 45 leader behaviors were developed specifically for this study based on 19 categories of leader behavior described by Yukl (1981). Items specific to the BCE were developed from the personal experience of the first author and informal conversations with other experienced CE officers. Within each behavior category, BCE actions were chosen that would be visible to or known by the wing and base commanders. Also, an attempt was made to include actions expected to be rated high, neutral, and low in their indication of leadership by the BCE. Respondents were asked to indicate the extent to which each behavior would reflect leadership by the BCE. A seven-point Likert scale was used, ranging from "very poor" (assigned a scale value of -3) through "not related" (0) to "very good" (+3).

Results

Of the 251 surveys mailed, a total of 160 were returned, yielding an overall response rate of 63.7%. Of this total, 42 identified themselves as wing commanders, 51 were base commanders, and 64 indicated they were BCEs; three surveys were missing demographic information. The number of responses were fairly representative of different base sizes and major air commands.

The leader behavior items were categorized into groups representing good leadership (mean responses of 1.5 or greater for at least two groups of respondents) or poor leadership (mean responses of -1.5 or less for at least two groups), with the remainder being labelled "not related" to leadership. Table 1 shows the behavior items rated as indicating good leadership quality, listed in order of the mean ratings by the wing commanders. The far right hand column indicates where significant differences in ratings between the respondent groups were observed (using the t-test).

Table 1
Behaviors Rated as Good Leadership Quality

BCE Behavior Item	Mean Rating [1] Wing	Base	BCE	T-TEST [2]
(The BCE) enforces strict adherence to AFR 35-10.	2.6	2.5	2.5	
. . . personally visits most CE job sites.	2.6	2.4	1.6	b,c
initiates meetings to brief commanders.	2.5	2.5	2.2	
frequently invites commanders to visit CE area.	2.5	2.3	2.1	b
publicizes CE activities in the base paper.	2.5	2.2	2.3	
anticipates commanders' desires and then acts.	2.4	2.3	2.2	
strongly presents CE position at meetings.	2.3	2.3	2.5	
frequently meets socially with peers on staff.	2.2	2.1	1.8	b
ensures special projects receive close attention.	2.1	2.1	2.2	
lives on base.	2.1	2.2	1.4	b,c
delegates decision making authority.	2.0	1.8	2.3	
uses formal military titles and courtesies.	2.0	1.9	1.8	
uses informal meetings to transfer information.	1.8	1.4	1.7	
consults with CE staff before making decisions.	1.8	1.2	1.9	a,c
ensures senior CE officers rate junior officers.	1.7	1.6	0.8	b,c
frequently wears the fatigue uniform to work.	1.6	1.5	0.5	b,c

[1] Ratings: 3 = Very Good; 2 = Good; 1 = Mildly Good
[2] T-Test Results: a = Sig. diff. between wing and base commanders ($p<.05$)
b = Sig. diff. between wing commander and BCE "
c = Sig. diff. between base commander and BCE "

From these items, it appears that BCE leadership behavior involves enforcing high standards, taking action, initiating communication, setting a good example, and taking an active interest in the CE work force. Several items rated high by the commanders were seen as less good (or perhaps less relevant) to the BCEs, including visiting most job sites, living on base, and wearing fatigues. On the other hand, the BCEs emphasized delegation and consultation with their staff more than did the commanders.

Table 2 lists those behaviors rated as poor in leadership quality, again shown in order of the mean ratings by the group of wing commanders. Significant differences between the groups (using the t-test) are indicated in the far right hand column.

Table 2
Behaviors Rated as Poor Leadership Quality

BCE Behavior Item	Mean Rating [1] Wing	Base	BCE	T-TEST [2]
(The BCE) permits relaxed appearance standards.	-2.6	-2.1	-2.3	
. . . keeps CE activities out of the base paper.	-2.4	-1.9	-2.3	
seldom inspects CE personnel.	-2.4	-1.9	-1.6	b
seldom attends base-level functions.	-2.2	-1.9	-2.3	
lives off base.	-2.0	-1.7	-1.1	b
meets with staff members only in formal meetings.	-1.8	-1.5	-1.6	

[1] Ratings: -3 = Very Poor; -2 = Poor; -1 = Mildly Poor
[2] T-Test Results: a = Sig. diff. between wing and base commanders ($p<.05$)
b = Sig. diff. between wing commander and BCE "
c = Sig. diff. between base commander and BCE "

Not surprisingly, the results for poor leadership behaviors validate the previous findings for good leadership since many of the opposite behaviors are seen in the list in Table 2. Poor BCE leadership behaviors can be described as passive and uninvolved. Actions (or inaction) leading to low standards and low involvement with the base were rated as poor. BCEs, however, put less emphasis on behaviors they may have considered irrelevant or unrealistic, such as living on base and personally inspecting their workforce.

These results were further validated by written comments from the respondents regarding BCE actions most damaging to good leadership. Commanders cited actions such as non-responsiveness, inaction, defensiveness, and poor information gathering; their emphasis appeared to be on the BCE getting the job done for the commander. BCEs most frequently listed failure to delegate, poor communication, inconsistency, and favoritism, reflecting an emphasis on people oriented activities within the unit.

Discussion

The results showed that superiors (wing and base commanders) do indeed perceive leadership differently than the leaders (BCEs). For the commanders, the effect of BCE actions on overall mission performance was the greatest influence on the perception of leadership. Response patterns and written comments seemed to indicate great interest by the commanders in the initiation of upward communication and support of the military structure and procedures by the BCEs. The BCEs highlighted the effects of their actions on CE personnel as most important. They consistently emphasized lateral and downward communication, delegation of authority, and consistency. These two perspectives could conceivably produce great conflict.

At the same time, a great deal of consistency among groups must be recognized from these data. BCEs can, for the most part, satisfy both their own expectations of leadership and those of their superiors with a single set of consistent behaviors. Yet they must continually recognize the need to strike a balance between satisfying their own people and satisfying the boss.

Knowing how commanders perceive leadership should help BCEs understand how they are evaluated by their superiors. These results should provide a basis for initiating communication with commanding officers. In some cases, differing perceptions need to be resolved; in other cases, communication is needed because both groups agreed that it was a vital element in getting the job done.

In terms of further research, a more complete picture would be generated by gathering subordinate perceptions of leadership to compare with those of the superiors and incumbent leaders. Other potential rating groups which may have different perspectives on leadership by the BCE would be those in the CE command hierarchy and those in other base agencies that are essentially "customers" of CE. The comparison and combination of data from multiple sources such as these would produce a more complete picture of the leadership requirements of the BCE or any leader.

References

Calder, B.J. (1977). An attribution theory of leadership. In Staw, B.M., & Salancik, G.R. (Eds.) New directions in organizational behavior. Chicago: St. Clair Press.

Haenisch, J.P. (1984). Attribution of base civil engineering leadership by wing and base commanders. Unpublished master's thesis, Air Force Institute of Technology, Dayton, OH.

House, R.J., & Baetz, M.L. (1979). Leadership: Some empirical generalizations and new research directions. Research in Organizational Behavior, 1, 341-423.

McKnight, R.D., & Parker, G.P. (1983). Development of an organizational effectiveness model for base level civil engineering organizations. Unpublished master's thesis, Air Force Institute of Technology, Dayton, OH.

Pfeffer, J. (1977). The ambiguity of leadership. Academy of Management Review, 2, 104-112.

Rush, M.C., Thomas, J.C., & Lord, R.C. (1977). Implicit leadership theory: A potential threat to the internal validity of leader behavior questionnaires. Organizational Behavior and Human Performance, 20, 93-110.

The Relationship Between Leadership and Job Satisfaction/ Career Commitment of Air Force Personnel

Robert E. Niebuhr
Auburn University

Mickey R. Dansby
Leadership and Management Development Center (LMDC)

Abstract

Using the Air Force's Organizational Assessment Package (OAP) data base, a positive correlation was found between leadership behavior and both job satisfaction and career commitment intentions. Relationships involving supportive leadership behaviors showed stronger positive correlations than relationships involving instrumental (structuring) leadership behaviors. Organizational tenure moderated the relationships between leadership behaviors and career commitment intentions--individuals who had been in the service a shorter time (low tenure) had much lower perceptions of leadership supportive behaviors than did high tenure individuals. This finding is particularly noteworthy given the high attrition rate of first term personnel. The lower level of meaningful leader-follower exchanges for low tenure individuals and the corresponding low retention of these individuals led to several specific recommendations regarding military views of leadership and leadership training.

Introduction

The concept of leadership is one of the most researched behavioral variables in both military and non-military writings. Bass (1981) cites more than 5,000 references in his review of leadership research. Even so, many suggest we have yet to adequately define leadership in scientific and practical terms, with some questioning the very validity of the concept as a scientific construct (Washbush, 1984; Yoos, 1984). Within military circles, this lack of clarity can be especially frustrating, since the military services place such great emphasis on the teaching and practice of leadership (Washbush, 1984). Perhaps a contributing factor to the confusion is the overemphasis in military leadership writings on the directive aspects of leadership (Van Fleet & Yukl, 1985), to the neglect of the supportive, interpersonal aspects of leadership.

Numerous studies have suggested that these supportive leadership behaviors correlate positively with job satisfaction (Bass, 1981). A smaller number of studies suggest supportive leadership behaviors correlate negatively with job turnover (Fleishman & Harris, 1962; Graen, Liden & Hoel, 1982). On the other hand, the relationship between instrumental (directive) leadership behaviors and job satisfaction is not clear, though some suggest instrumental leadership behaviors correlate positively with turnover rate (Fleishman & Harris, 1962; Sarason, 1981).

Other research has shown a positive correlation between tenure with the organization and job satisfaction (Spencer & Steers, 1981) and a negative correlation between tenure and turnover (Mobley, 1977). Hall (1976)

suggests new employees (low tenure) are preoccupied with establishing their identity in the organization. Consequently, learning how to work with one's superiors becomes extremely important at lower tenure levels. At this stage, one might expect a superior's leadership behavior to have a strong influence on job satisfaction and turnover. As the employee's tenure increases, other variables (level of advancement, outside opportunities, etc.) may become more important determinants of satisfaction and turnover.

Based on this previous research, the authors proposed two hypotheses concerning the relationship between leadership behaviors and job satisfaction/career commitment:

(1) A positive correlation exists between leadership behaviors and job satisfaction/career commitment; supportive leadership behaviors should be more positively correlated with satisfaction and commitment than are instrumental behaviors.

(2) Organizational tenure will moderate the relationship between leadership behaviors and satisfaction/commitment; leadership variables will correlate more strongly with satisfaction/commitment at lower levels of tenure.

Method

Data for the present study came from the LMDC Organizational Assessment Package (OAP) data base of almost 300,000 survey responses collected in LMDC management consultation visits to Air Force organizations. The surveys were administered from October 81 to July 85 as a census of organizations (usually wing size or equivalent) at 65 bases or sites where LMDC was invited to consult. Surveys were administered in group sessions and respondents were assured of individual anonymity of responses. Only LMDC consultants handled completed survey response sheets. Respondents for the present study were a 31,980-person segment of the data base, chosen to provide a representative sample of functional areas (operations, maintenance, and resources) and personnel categories (officer, enlisted, and civilian).

The OAP provides an overall leadership scale of nineteen items. Six of these items were chosen on the basis of their content to form a measure of instrumental leadership; four others were chosen to provide a measure of supportive leadership behavior. The Job Satisfaction OAP factor consists of seven items and has been shown to be reliable (Short, 1985). Career commitment was measured using a single OAP item, and organization tenure was derived from the demographic section of the OAP.

Results/Discussion

Descriptive statistics, scale reliabilities (where appropriate), and Pearson product-moment correlations among the variables are presented in Table 1. The results support the first hypothesis, which predicted positive correlations between leadership behavior and both job satisfaction and career commitment. Although the leadership subscales are highly intercorrelated, the stronger relationships between supportive leadership behaviors and both job satisfaction and career commitment are as hypothesized.

Table 1

Descriptive Statistics, Reliabilities, and Intercorrelations Among the Study Variables

	M	SD	1	2	3	4	5
1. Overall Leadership	4.79	1.46	(.97)				
2. Instrumental Leadership	4.78	1.54	.96	(.93)			
3. Supportive Leadership	4.75	1.47	.90	.83	(.83)		
4. Job Satisfaction	5.07	1.16	.50	.47	.54	(.81)	
5. Career Commitment	3.74	1.31	.15	.15	.20	.31	
6. Organization Tenure	4.88	1.87	.01	.03	.08	.10	.48

Notes: N = 31,980
Reliability coefficients are in parentheses in the diagonal. All correlation coefficients above .02 are significant at the .001 level.

Multiple regression analysis was used to evaluate the second hypothesis (see Table 2).

Table 2

Results of Moderated Regression Analysis for Leadership Measures, Job Satisfaction, and Career Commitment

Dependent Variable	Regression Variables	Interaction Effect R^2	ΔR^2	F
Job Satisfaction	OL, OT, OL x OT	.252	.000	0.07
Job Satisfaction	IL, OT, IL x OT	.229	.000	0.03
Job Satisfaction	SL, OT, SL x OT	.298	.000	0.39
Career Commitment	OL, OT, OL x OT	.256	.001	28.62*
Career Commitment	IL, OT, IL x OT	.254	.001	14.57*
Career Commitment	SL, OT, SL x OT	.264	.001	29.99*

Notes: df = 1, 21811; OL - Overall Leadership; IL = Instrumental Leadership; SL = Supportive Leadership; OT = Organization Tenure; $^*p \leq .001$

For moderated regression analysis, the hierarchial inclusion method was used for decomposition of the explained sum of squares into components attributable to each independent variable and the interaction term. The multiplicative interaction term was always entered into the regression in the last step. Therefore, the F test associated with the interaction term indicates the significance of the incremental contribution of this term to the explained variance in the dependent variable after the effects of all other independent variables have been removed.

The interaction of leadership behaviors and organization tenure added significantly to the explained variance in career commitment but did not with respect to job satisfaction. To evaluate these interactions more carefully the data were split into subgroups of low and high organization tenure. Table 3 presents the subgroup means of each of the study variables.

Table 3

Analysis of Variable Means by Low and High Organization Tenure Subgroups

	Low Tenure	High Tenure
Job Satisfaction	4.94	5.23
Career Commitment	3.19	4.49
Overall Leadership	4.73	4.87
Instrumental Leadership	4.70	4.89
Supportive Leadership	4.62	4.93

Note: *t*-tests on the differences in the subgroup means indicate significance for all comparisons at the .001 level.

Using Vecchio's (1985) terminology, these results might suggest a *repulsion* causal model is a stronger rationale for this population's attitudes than an *attraction* causal model. The correlation between leadership and career commitment intentions was both positive and significant (in Vecchio's attraction causal study the coefficient was slightly negative and non-significant), although it did not exceed the level of the correlation between leadership and job satisfaction (as in the study by Graen, Liden, and Hoel, 1982). Substantial job irritants affect all personnel, causing dissatisfaction and reduced commitment. For low tenure individuals, effective supportive-oriented leader behaviors should help dissipate these negative feelings and reduce the individual's propensity to leave. However, the data in Table 3 indicate both perceived supportive leadership behavior and career commitment intentions were lower for low tenure than for high tenure individuals. Apparently, the low tenure individuals do not perceive their leaders are providing the type of support required to help them deal with job irritants. This repulsion causal rationale receives support from the study by Sarason (1981), reviewed earlier, which found higher attrition rates in Marine Corps platoons where leaders did not provide sufficient counseling and consideration behavior to new recruits. For high tenure individuals, perhaps the socialization process has made them more sure of themselves and more comfortable with their environment. Additionally, the vested tenure and shorter time to retirement would tend to increase their career commitment.

Recommendations

The implications of this study suggest several recommendations for both military policy changes and research directions:

(1) Military personnel who function in supervisory roles need to be provided with a clear image of the concept of leadership and its impact on follower attitudes and behaviors.

(2) Military training of supervisory personnel needs to be reexamined with respect to content and application of leadership principles. Particular emphasis needs to be placed on those supervisory personnel involved with low organization tenure individuals.

(3) Additional research and policy changes need to be accomplished which would isolate and reduce the primary irritants which serve to decrease career commitment.

References

Bass, B.M. (1981). Stogdill's handbook of leadership: A survey of theory and research. New York: Free Press.

Fleishman, E.A. & Harris, E.F. (1962). Patterns of leadership behavior related to employee grievances and turnover. Personnel Psychology, 15, 43-56.

Graen, G.B., Liden, R.C., & Hoel, W. (1982). Role of Leadership in the employee withdrawal process. Journal of Applied Psychology, 67:6, 868-872.

Hall, D.T. (1976). Careers in organizations. Pacific Palisades, CA: Goodyear.

Mobley, W.H. (1977). Intermediate linkages in the relationship between job satisfaction and employee turnover. Journal of Applied Psychology, 62, 237-240.

Sarason, I.G. (1981). Psychological approaches to Marine Corps recruit attrition. In H. Sinaiko et al. (Eds.) Military personnel attrition and retention research (TR-10). Washington: Smitnsonian Institute.

Short, L.O. (1985). The United States Air Force Organizational Assessment Package (Report No. LMDC-TR-85-2). Maxwell AFB, AL: Leadership and Management Development Center.

Spencer, D.G. & Steers, R.M. (1981). Performance as a moderator of the job satisfaction-turnover relationship. Journal of Applied Psychology, 66, 511-514.

Van Fleet, D.D. & Yukl, G.A. (1985). Military-leader behavior: Past, present, and future. In J.G. Hunt & J.D. Blair (Eds.), Leadership on the future battlefield. McLean, VA: Pergamon-Brassey.

Vecchio, R.P. (1985). Predicting employee turnover from leader-member exchange: A failure to replicate. Academy of Management Journal, 28:2, 478-485.

Washbush, J.B. (1984). Leadership instruction and officer candidates: Some unholy thoughts. In Proceedings of the ninth symposium: Psychology in the Department of Defense. Colorado Springs: United States Air Force Academy.

Yoos, C.J., II, (1984). There is no such thing as leadership! In Proceedings of the ninth symposium: Psychology in the Department of Defense. Colorado Springs: United States Air Force Academy.

Pilot Selection and Performance Evaluation: A New Look at an Old Problem

Robert L. Helmreich

The University of Texas at Austin

Abstract

Limitations on the validity of pilot selection strategies are discussed. Major problems are caused by clinging to an obsolete sterotype of the pilot as a single figure facing the elements in isolation and by limiting the performance criterion to behavior during training. It is argued that this leads to an incorrect underemphasis on the role of personality factors in long term performance. Performance criteria should be obtained in an operational environment with stress on <u>crew</u> as well as individual behavior. The implications of changing approaches to evaluation on selection are discussed.

Background

The selection of the best qualified candidates as pilots for high performance military aircraft was a major concern of American psychologists during World War II and has continued to be a significant research issue in the ensuing forty years (Melton, 1947; Helmreich, 1982). It will be argued here that, although the question has broadened to include the selection of crewmembers for spacecraft and for sophisticated civilian jet transports, research in 1986 continues to suffer from the same major failing that limited its utility in the piston era, the use of inappropriate performance criteria. The central thesis is that deficiencies in the criterion lead to overemphasis on some predictors and the neglect of others. Let us begin, however, by examining the attributes of an outstanding pilot.

The Red Baron Revisited: Prototype or Dinosaur?

The prototypical pilot continues to be drawn from the era of white scarves and open cockpits: a man with lightning reflexes, high intelligence, extraordinary courage, X-ray vision, and a legendary capacity for alcohol and sleep-loss, an individual who can be characterized as having "The Right Stuff" (Wolfe, 1979) or being a "Macho Pilot" (Helmreich, 1979). The model, then, is the fighter pilot, singlehandedly challenging the equipment, the elements, and the enemy. Perhaps the best living exemplar of all this is General Charles Yeager whose autobiographical candor gives us a good picture of the man behind the image (Yeager & Janos, 1985). In the less florid language of experimental psychology, we could describe the best of this genre as having

extremely high aptitude and motivation for the task at hand. This brings us to the central question. How do we know when we have selected an individual with the requisite qualities?

Success in Training as the Criterion of Selection.

Arthur W. Melton (1947), in his review of pilot selection results from World War II research, noted that the use of training success as the selection criterion was a notable limitation on the validity of the research. This remains true today in both military and domestic and foreign airline pilot selection (Goeters, 1980; Helmreich, Hackman, & Foushee, 1986). The multiple pressures from organizations and individuals to restrict operational evaluation have of individual and crew performance been discussed in detail elsewhere (Helmreich, Hackman, & Foushee, 1986).

The ultimate performance criterion, of course, is how the individual performs over time in the operational setting. In the case of the single pilot, non-automated aircraft, many of the operational requirements may be highly similar to those encountered during initial training. However, the motivation to sustain performance under a variety of conditions may not be assessable during training. In the case of multi-crew aircraft, overall performance is as much a result of the ability of the <u>team</u> to work effectively than of individual proficiency. NASA research has demonstrated conclusively that the majority of jet transport accidents and incidents result from failures in crew coordination rather than mechanical problems or deficiencies in the technical competence of crewmembers (Cooper, White, & Lauber, 1979). This is particularly important because the majority of pilots selected now and in the future will operate in a multi-crew environment.

A recent study by Helmreich, Sawin, and Carsrud (in press) demonstrates a major weakness of the use of initial training performance as the selection criterion. In this study personality and motivational factors measured prior to employment proved to be good predictors of job performance. However, this prediction was obtained only after the individuals had been out of training and on the job for more than three months. The predictors were unrelated to performance in training and after initial release to the workforce. The authors refer to this as a "honeymoon effect", arguing that most candidates for a desired position will exert maximum effort during an initial "honeymoon" period and that only in the longer term will underlying dispositions become significant determinants of behavior.

The Neglected Role of Personality.

Melton (1947) stressed the conceptual importance of personality factors in pilot performance. In practice, however, personality factors have played a very limited role in civil and military selection and that role has been primarily in screening out psychopathology rather than selecting <u>for</u> desired attributes

(Helmreich, 1983). Indeed, a review of twenty-five years of pilot selection research in the U. S. Navy concluded that personality measures have not proved to play any significant role in predicting performance in training attrition (Griffin & Mosko, 1977).

It is certainly possible that limitations in the performance criterion, including the "honeymoon effect" during training, rather than deficiencies in the conceptualization and measurement of personality account for the paucity of positive findings. There are also positive results showing personality factors as predictive of flying performance. Helmreich (1982) used Check Airmen's ratings of line performance as the dependent measure and found that trait constellations of instrumentality and expressiveness as well as components of achievement motivation were significantly related to this operational criterion. In a massive study of physiological and psychological reactions to short-haul airline operations, NASA researchers found that the same instrumental and expressive personality dimensions were significant predictors of objective and subjective measures of operational fatigue, with crewmembers high on both dimensions showing the most positive responses (Foushee, 1986). Similarly, researchers at the Deutsche Forschungs und Versuchanstalt fur Luft und Raumfahrt (DFVLR) in Germany successfully employ personality measures in the selection of ab initio candidates for airline pilot positions, although the performance criteria do not extend beyond a two year training program (Goeters, 1980). In any event, it would seem premature to conclude that personality factors cannot account for significant amounts of variance in flightcrew performance.

Problems and Promise in Performance Evaluation.

While the evaluation of individual, technical performance both in simulators and aircraft has become quite sophisticated, the assessment of crew performance and the managerial aspects of flight has lagged far behind. Ivan Steiner (1974) was moved a decade ago to ask "Whatever happened to the group in social psychology?", noting sadly that research on group processes and performance had almost disappeared from the literature - and the situation has not changed dramatically in the intervening period! Part of the avoidance of group studies undoubtedly comes from the methodological difficulties encountered in dealing with the multi-variate complexities of human interaction and partitioning behavioral variance among individuals and the group.

The picture is not entirely dark, however. As Foushee (1984) has noted, the modern aircraft simulator is not only an effective training device, but also an outstanding setting for conducting research on crew interaction and performance. With advanced simulation facilities, it is possible for crews to fly complete missions, including ground and cabin communications, using a standardized scenario with problems requiring both managerial and technical solutions. This provides a research setting with unparalleled control and realism. Recognizing the

research value of this paradigm, NASA has recently completed a facility at the Ames Research Center which has two jet transport simulators and air control simulation capabilities along with sophisticated recording and data reduction equipment. This facility will be completely dedicated to research on the human performance component of multi-crew jet operations. In tandem with the development of this facility, a number of researchers are facing the challenge of trying to develop and refine methodologies to analyze and understand the dynamics of _crew_ behavior.

Both the military and the airlines have also recognized the enormous utility of full mission simulation for training and are enthusiastically implementing it under the name of Mission Oriented Flight Training (MOST) in the military and Line Oriented Flight Training (LOFT) in civilian transport. What is lacking is better evaluation tools for instructors and Check Airmen and the exploitation of the _research_ potential of this type of training.

An Optimistic Scenario.

While the research data are still sparce and the required methodologies underdeveloped, it seems highly likely that, with continued research support, much more valid criteria of both individual and full crew performance will evolve. When such criteria are available, they can be employed to evaluate experienced pilots in both simulator and line operations. Given such validated criteria, it will be possible to re-examine the entire process of pilot selection and to determine which predictors, including personality measures, are associated with optimal _line_ performance.

There are several other areas which will benefit from the refinement of performance evaluation. These include defining more clearly the capabilities and limitations of selection and training and assessing the effectiveness of various training techniques, especially training in crew coordination and cockpit resource management.

In the long term, an especially important contribution of better performance evaluation will be to our understanding of the performance impact of cockpit automation and the operations in the "glass cockpit" environment - including specification of the type of individual best suited to operate effectively in such an environment.

References

Cooper, G.E., White, M.D. & Lauber, J.K. (Eds.), (1979) _Resource Management on the flight deck_. NASA Report No. CP-2120.

Foushee, H.C. (1984) Dyads and triads at 35,000 feet: Factors affecting group process and aircrew performance. _American Psychologist_, _39_, 886-893.

Foushee, H.C. (1986) NASA Crew Fatigue Project, personal communication.

Goeters, K. M. (1980) Selection of personnel for complex operations as demonstrated by pilot selection. Paper presented to the Ispra-Course "Training of operating personnel for technologies with hazard potential: Theory and practice. Ispra, Italy.

Griffin, G.R. & Mosko, J.D. (1977) Naval aviation attrition 1950-1976: Implications for the development of future research and evaluation. Pensacola, Florida: Naval Aerospace Medical Research Laboratory.

Helmreich, R.L. (1979) Social psychology on the flight deck. In Cooper, G.E., White, M.D. & Lauber, J.K. (Eds.), (1979) Resource Management on the flight deck. NASA Report No. CP-2120.

Helmreich, R.L. (1982) Pilot selection and training. Paper presented at the American Psychological Association, Annual Meeting, Washington, D.C.

Helmreich, R.L. (1983) What changes and what endures: The capabilities and limitations of training and selection. In N. Johnson (Ed.) Proceedings of the Aer Lingus/Irish Airline Pilots Association Flight Symposium, Dublin, Ireland.

Helmreich, R.L., Sawin, L.L. & Carsrud, A.L. (in press) The honeymoon effect in job performance: Delayed predictive power of achievement motivation. Journal of Applied Psychology.

Helmreich, R.L., Hackman, J.R. & Foushee, H.C. (1986) Evaluating flightcrew performance: Policy, pressures, pitfalls, and promise. NASA/University of Texas Technical Report 86-1. Austin, Texas.

Melton, A.W. (1947) Army Air Forces aviation psychology program: Report No. 4: Apparatus tests. Washington, D.C.: Defense Documentation Center.

Steiner, I.D. (1974) Whatever happened to the group in social psychology? Journal of Experimental Social psychology, 10, 94-108.

Wolfe, T. (1979) The right stuff, New York: Bantam.

Yeager, C. W. & Janos, L. (1985) Yeager, New York: Bantam.

The author's research reported here and preparation of the report were supported by NASA Cooperative Agreement NAG2-137, Robert L. Helmreich, Principal Investigator.

Considerations in Evaluation of Cockpit Resource Management Training

John Wilhelm

The University of Texas at Austin[1.]

Abstract

A short history of Cockpit Resource Management Training is given. "Macro" evaluation strategies as well as strategies for evaluation of individual CRM programs are discussed. Changes in pilot and crew performance on the line are argued as the best indicators of CRM effectiveness.

INTRODUCTION: Considerable work in the area of Cockpit Resource Management Training has been accomplished in recent years. Major commercial airlines have developed courses around the topic and are including it as a part of their initial and recurrent training requirements for pilots. Smaller 121 carriers and Commuters who have been unable to develop their own programs are being courted by both major airlines and independent corporations who sell training for a profit. Corporate pilots are offered versions of the training at major Corporate training centers and independent companies. International. "Militarized" versions of this training are currently being utilized with some crews of C-130 and C-141 and courseware is currently being developed and adapted for C-5 crews. I would like to describe a short history of its development, then turn to a discussion of our thinking about evaluation of the CRM concept and programs.

IDENTIFYING THE PROBLEMS: In 1973, NASA began a "human factors in aviation" research program. Based on structured interviews with pilots the areas of crew training, cockpit management and leadership emerged as frequent sources of concern. Lauber (1979) describes this project and the analysis of jet transport accidents which proceeded in tandem with the interviews. His group of researchers found more than 60 accidents in the 1968-1976 time frame in which "resource management" was a contributing factor. Additional evidence of the validity of these problems began to show up in a simulator study by Ruffell Smith (1979) where pilots flying a scenario designed to exercise CRM skills made errors, some of which caused simulator "crashes". The NASA researchers also acknowledged that for the kinds of analyses we "statistical types" need, a larger database was needed. In 1975, NASA set up the Aviation Safety Reporting System which encourages pilots to report "incidents" in their own words. In return, the pilot reporting the incident is given immunity from prosecution and assured confidentiality. Currently there are more than 50,000 incidents of all types in this invaluable resource.

DEVELOPMENT OF TRAINING PROGRAMS: Although different names are applied by different organizations, a core curriculum has developed. Murphy (1980) conducted a detailed analysis of a sample of incidents from ASRS. He developed a conceptual scheme for describing the causes of the incidents and was able to get a distribution of accidents falling into each category. From these analyses the basic curriculum evolved. Generally, the training consists of exercises to improve social and communication skills, leadership and management skills and planning, problem solving and decision making skills. These are not the traditional "stick and rudder" exercises normally associated with pilot training.

LOFT: A couple of other developments have also helped the stampede gain momentum. The early deliberations about types of training which would affect the accident and incident statistics gave birth to the LOFT (Line Oriented Flight Training) concept. In this training, a full crew is given a training mission to fly in the simulator. It is flown without interruption and in real time. LOFT gives the trainees a chance to interact and practice concepts which have been learned in CRM. Sessions are usually videotaped, with the tapes used for extensive debriefing by a trained LOFT instructor. The combination of academic CRM with LOFT for hands-on practice and videotaped replays of crew interaction is a powerful training tool. (Lauber & Foushee, 1981)

Based on encouraging results from the early applications of this training at Northwest Orient, in May, 1978, the FAA issued an advisory circular (AC 120-35) to the FAR 121 to allow LOFT to be utilized in airline recurrent training programs. Additionally, to encourage its use, if a carrier combines CRM with LOFT, airlines may request an exemption to conduct pilot training annually instead of twice yearly. This action gives economic impetus to carriers to change their training procedures--annulized training should save more money to carriers than the additional training costs associated with LOFT.

EVALUATION POSSIBILITIES: Thus, in a remarkably short period of time, the concepts have not only been accepted but also have been implemented across a wide range of aviation settings. The gut feeling of most observers is that we are actually affecting safety in the skies--But are we? Have we really affected the number of accidents or incidents out there, or dropped the percentage of accidents or incidents which relate to "pilot error" as defined as failures in crew coordination? As the statistics were the main impetus for developing the training concepts, they should be affected by our training efforts. A chilling possibility is that we are all victims of group-think: that we have all agreed on the problem and have designed curricula to address it, have congratulated ourselves that the problem is solved without verifying the correctness of this

assumption. Most programs have feedback from participants about which parts of the program were liked and disliked and overall reactions to the program, but is this the proper evaluation of CRM?

We think that evaluation should proceed at various levels simultaneously. At the macro level, does the program have face validity--i.e. does it affect accidents and incidents positively? Additionally, we see the model which seems to underlies CRM as critical for evaluation. Academic instruction is emphasized by exercises and group discussions. LOFT, if it is a part of the training, is a practice session used not only to practice traditional stick and rudder operations, but also to practice CRM concepts in "coached" environment. This training is then carried into the cockpit where the new attitudes and communication skills should produce a more relaxed, coordinated and paced atmosphere. The crew will exercise improved decision making and resource management. These improvements, in the long run, will affect accident and incident statistics. The training is periodically reinforced by recurrent training and additional simulator session. The training works through improved attitudes and resource management skills, which can and should be evaluated as well as accident and incident statistics. Thus components to be evaluated include attitude change, specific behavioral objectives (SBO's) during training, performance in the simulator and on the line as well as accidents and incidents.

The question of evaluation is complex. There are many programs, all with a similar core curriculum, but with differing methods and emphasis. Some CRM programs are coupled with LOFT, others are not. It is a given that some programs will be more effective than others. All evaluation approaches take the general form of Pre-measures, training and then post-training measures, and a comparison of the two. What the measures are and what can be realistically be collected given the constraints dictated by regulatory agencies, companies, and individual pilots and pilot groups are being examined currently.

MACRO EVALUATION: At the "macro" level, we all assume that the programs taken as a whole will affect the number of accidents and incidents. Here we are evaluating the concept of CRM itself in terms of its expressed purpose which is to reduce accidents and incidents. Historically comparing accidents and incidents before and after CRM may be possible in the aggregated long-term, but the effects will be long-term and slow to make themselves known. There is a very large pool of pilots out there! This kind of evaluation may be possible using ASRS data. Demonstrating effects will also require control over other historical trends that are occurring and which also affect accidents and incidents. These trends include increasing automation even within A/C type, new types of A/C, increasing numbers of pilots, changes in the ratio of military vs civilian pilots, weather variations year-to-year, increasingly smaller crews, etc. A more effective approach would be to follow a large sample of pilots in a given A/C type recording accident and

incident rates before and at points as CRM makes its way through the population. It is doubtful whether this approach could work within a single carrier as the accident and incident rates are so low to begin with.

EVALUATION OF INDIVIDUAL PROGRAMS: At the individual level, a pilot's "performance" should improve as a consequence of CRM. Helmreich (1986) has discussed some of the many dimensions of performance evaluation in his paper. As the pilot proceeds through training, a set of specific behavioral objectives (SBO's) should be more or less achieved and attitudes should begin to change in a pre-defined direction. Each CRM program should have its own SBO's. Given the difficult task of training nebulous communication, leadership and management skills, definition of SBO's will not be easy. To the extent that the SBO's are attained by pilot classes, the training was successful. At the completion of the training, feedback should be elicited within each program to help program designers refine the individual program components.

Line and LOFT performance--Individual: Whether accomplishment of the SBO's affects actual line performance is at the heart of the question. Pilot performance is regularly evaluated by Check Airmen both in simulators and on the line. If groups of Check Airmen are brought together to discuss dimensions of evaluation, agreement can be reached as to what dimensions relate to good performance. Then ratings can be applied to required checks and systematically collected. Thus a baseline level of individual effectiveness can be established against which CRM effects can be measured. As Helmreich has stated, performance out on the line is the truest indicator of the success of CRM. Performance in the simulator is probably the next best indicator of CRM effects.

Crew performance: A current hot topic is evaluation of crew performance. CRM and LOFT are based on the crew concept. Our traditional evaluations of individuals are possible within this framework, but a challenge is to examine performance of crews both before and after CRM. These differences will likely be the most compelling evidence for validation of CRM.

Changes in Attitudes: A basic change in the attitudes pilots have about their role is an objective of training. Helmreich et. al (1985) demonstrated that cockpit management attitudes relate to Check Airmen ratings of line performance. Thus a change of attitudes in the positive direction can be seen an evidence that CRM is working.

Training Environment Considerations: Each CRM program should be evaluated separately. Each is different in format and focus and will have different SBO's. Some include LOFT, some do not. Additionally, comparison across programs will be difficult as the environment surrounding each application is so different. Military and commercial programs differ in focus and objective, as do Union vs Non-Union environments within 121 carriers.

Measures which would ideally be individual based may have to be considered for de-identified groups only because of environmental and ethical concerns.

PROPRIETARY CONSIDERATIONS: As these individual program evaluations are conducted, consideration must be made of the fact that many of these programs are proprietary. Evaluation documents will possibly be used for internal consumption only. Some type of Federal support to assist the efforts (FAA, NASA or DOD) should be extended so results may become public.

REFERENCES

Helmreich, R.L., Foushee, H.C., Benson, R. & Russini, W. (1985). Cockpit resource management: Exploring the attitude-performance linkage. In Proceedings of the Third Aviation Psychology Symposium, Columbus Ohio: The Ohio State University.

Helmreich, R.L. (1986). Pilot selection and performance evaluation: A new look at an old problem. In Proceedings of Psychology in the Department of Defense Ninth Symposium, Colorado Springs, Colorado: US Air Force Academy.

Lauber, J. K. (1979). Resource management on the flight deck: Background and statement of the problem. In G. E. Cooper, M. D. White, and J. K. Lauber (Eds.), Resource management on the flight deck. NASA CP 2120. Moffett Field, CA.

Lauber, J. K. & Foushee, H. C. (1981). Guidelines for Line-Oriented Flight Training. Vol. 1, NASA CP-2184. Moffett Field, CA.

Murphy, M. (1980). Analysis of eighty-four commercial aviation incidents: Implications for a resource management approach to crew training. In Proceedings of the Annual Reliability and Maintainability Symposium, IEEE.

Ruffell Smith, H. P. (1979). A simulator study of the interaction of pilot workload with errors, vigilance, and decisions. NASA TM 78482. Moffett Field, CA.

Footnotes:

1. The authors research reported here and preparation of the report were supported by NASA Cooperative Agreement NAG 2-137, Robert L. Helmreich, Principal Investigator.

The Effects of Aircrew Member Personality on Interaction and Performance

Frederick M. Siem
The University of Texas at Austin1

Abstract

Eighty pilots whose responses to a personality questionnaire identified them as having characteristics associated with either superior or inferior flightdeck performance were selected to work in two-person crews operating a microcomputer-based flight simulator program. Raters viewed tapes of the simulator sessions and evaluated performance and communication styles. Analyses indicated that crews in which the pilot was highly competitive and low in an expressive, interpersonal orientation tended to communicate more with their co-pilots and that higher levels of interaction were associated with poorer performance, in contrast with studies of airline pilots (e.g. Foushee & Manos, 1981).

Introduction

The observation that 65% of aviation accidents can be attributed to "pilot error" (Foushee, 1984) has focused much research activity on the contribution of pilot personality characteristics to performance on the flight deck. Helmreich (1982), for example, examined the relationships between personality traits, achievement motives and performance measures on a sample of airline pilots. He found that self-assertiveness and expressiveness were positively associated with both Check Airmen ratings and with simulator performance scores. In splitting the sample into thirds based on a criterion score based on a composite of the two performance measures, Helmreich found that the superior pilots were high on both self-assertiveness and expressiveness but low on interpersonal competitiveness.

While research to date provides compelling evidence for the contributions of pilot personality to performance, a less explored issue is how the characteristics of _individual_ members influence performance of flight _crews_. Also of particular interest is the question of how crew member characteristics affect interactions and communications. Ineffective crew coordination processes involving factors such as poor leadership, communication breakdowns, and personality clashes have been associated with a number of accidents attributed to pilot error (Foushee, 1984).

Concern with the process of crew interaction, and how this process is related to performance, has led investigators at the NASA Ames Research Center to conduct investigations of communications made by aircrew members participating in flight simulator studies. Foushee and Manos (1981) found that for 747 crews, operational errors were negatively correlated with crew

member observations about flight status, while acknowledgments were negatively correlated with systems errors and with overall errors, and commands were negatively associated with flying errors. The crews performing more poorly also demonstrated a tendency to communicate less and to express more response uncertainty, frustration, embarrassment and less agreement.

Consideration of the analyses by Foushee and Manos (1981) suggests that patterns of communication may mediate the influence of personality characteristics on flightcrew performance demonstrated by Helmreich (1982). Evidence for the impact of group member characteristics on communication and interaction comes from a series of studies by Ickes and colleagues (Ickes, 1981). In the paradigm common to these studies, two college students pre-selected on self-assertive and expressive characteristics participate in an unstructured 5-minute interaction. In each study, the highest level of interaction was found for dyads in which both members were high both in self-assertiveness and in an expressive, interpersonal orientation.

The present research was undertaken to replicate the results of the Ickes (1981) paradigm in a structured, task performance situation as an attempt to explore how aircrew member personality characteristics may affect communication style and performance. On the basis of the research discussed above, it was expected that expressive individuals will interact more, which should contribute to crew coordination and thus performance, whereas competitiveness should interefere with effective crew coordination.

Method

Subjects

80 male pilots took part in forty two-person crews performing a flight simulator task. For each crew, the pilot-in-command (PIC) had an Instrument Flight Rule (IFR) rating, while the person chosen to act as copilot always had a Visual Flight Rule (VFR) rating. The PIC's had a mean age of 41.9 years and an average of 2402 hours flying on both single and multi-engine aircraft; the average co-pilot age was 34.3 years, with a mean of 423 hours flying.

Procedure

A survey was sent to 3000 pilots in a Southwest metropolitan area. The first part of the questionnaire was comprised of items concerning demographic variables such as age, sex, education, pilot rating, and hours flown on different type of aircraft. The second part of the questionnaire included the Expressiveness scale of the Personal Attributes Questionnaire (Spence, Helmreich & Stapp, 1974), which measures a cluster of personality traits related to an interpersonal orientation. The third part of the questionnaire measured the enjoyment of interpersonal competition using a scale from the Work and Family Orientation Questionnaire (Helmreich & Spence, 1978).

Eighty pilots who scored above the median on one scale and below the median on the other scale were selected for the simulator study and assigned to one of four conditions. In one condition both pilots were highly expressive and low on competitiveness; in a second condition, both pilots were low on expressiveness and highly competitive. The other two conditions represented crews with a mix of characteristics: a competitive, low expressive PIC teamed with a low competitive, expressive co-pilot, or an expressive, low competitive PIC paired with a competitive, low expressive co-pilot. Thus, there were ten crews in each of the four cells of a 2 (pilot personality) X 2 (co-pilot personality) experimental design.

Pilots who met the criteria were contacted by phone and scheduled for a simulator session using a commercially available program for the IBM PC that emulates a Cessna 182. During the flight, normal radio communications were conducted with Air Traffic Control (ATC) over an intercom system. A second experimenter acting as ATC monitored the flight on a television set connected to a videotape deck that was used to record the control panel and window view seen by the pilot subjects operating the simulator. All verbal communications between the subjects and with Air Traffic Control were recorded onto the audio channels of the videotape.

During the flight, the PIC always flew the first segment and the co-pilot the second, so that each pilot an opportunity to become familiar with the handling characteristics of the simulator. For segments three and four, designed to increase the workload and the need for effective crew coordination, the PIC was allowed to distribute duties at his own discretion. At the end of the third segment, the crews were unable to tune in the navigational radio at their destination airport and the weather at the alternate destination listed in their flightplan was also "below minimums," so a third airport had to be selected. While the landing was in progress, the crew were instructed to abort and make a Missed Approach. After the missed approach was attempted, the simulation was ended.

Two sets of raters, blind to condition, viewed videotapes of the simulator sessions. One set of raters, all pilots themselves, rated performance on each flight segment along three dimensions: maintaining assigned altitude, correct heading, and overall performance. A second set of raters recorded interaction measures such as the number of turns each crew member took speaking. For each of the four flight segments, a five-minute section was transcribed and content-coded using categories such as observation, question, answer, agreement, disagreement, and command.

Results

Planned comparisons for mean task performance rating for each flight segment were conducted, contrasting the two conditions of main interest, the crew in which both members were

high expressive/low competitive vs. the low expressive/high competitive crews. Although the means for the two conditions were in the predicted direction for each flight segment, only the contrast for the first flight segment approached significance, $p < .06$. The low expressive/high competitive crews had the lowest rating for each segment compared to all other groups, but these differences were not statistically significant.

A series of 2 (pilot personality) X 2 (co-pilot personality) analyses of variance were conducted, the dependent variable being frequency of interaction by each pilot for each of the four flight segments. For flight segment two, during which the co-pilot was flying, low expressive/high competitive co-pilots and their PIC's interacted more (X = 80.4) than high expressive/low competitive co-pilots (X = 61.3)[$F(1,36) = 7.03$, $p < .05$]. For flight segment three, which involved a long, routine stretch ending in a problem situation requiring a decision about an alternate destination, the interactive effect of pilot and co-pilot personality types on frequency of interaction was significant [$F(1,36) = 4.10$, $p < .05$]. Examination of the mean frequency of interaction for co-pilots indicates that this interaction resulted from pilot/co-pilot crew members with _similar_ personality characteristics interacting more than _dissimilar_ personality types, with the high expressive/low competitive co-pilots manifesting the highest frequency of interaction. Also during this segment, contrary to the rest of the flight, high expressive/low competitive pilots interacted more (X = 42.5) than the competitive/low expressive pilots (X = 33.7).

Correlational analyses of the relationships between frequency of interaction with performance rating and crew member personality characteristics indicated that higher levels of interaction were positively associated with PIC competitiveness ($r = .27$, $p < .05$) and negatively correlated with PIC expressiveness ($r = -.27$, $p < .05$). Higher levels of interaction were also associated with lower performance ratings ($r = -.39$, $p < .01$).

In order to examine whether a particular type of communication was associated with the higher rates of interaction manifested by crews with low expressive/high competitive pilots, another series of analyses of variance were conducted for the frequency for each of twelve types of communication. The results of these analyses indicated that low expressive/high competitive pilots expressed more anger (X = 8.7 vs. X = 4.3) [$F(1,36) = 4.59$, $p < .05$] whereas the high expressive/low competitive co-pilots verbalized more agreements and acknowledgements (X = 19.3 vs. X = 12.8) [$F(1,36) = 4.02$, $p < .05$].

Discussion

As predicted, there were differences in interaction level and types of communication associated with crews composed of different personality types. Contrary to the data from

simulation studies with airline pilots, however (e.g. Foushee & Manos, 1981), higher rates of interaction were associated with lower performance ratings. Differences between the task and the samples used in the two studies may explain these contradictory findings. Foushee and Manos (1981) analyzed tapes involving professional pilots with multicrew experience performing a fairly structured task; in contrast, the present study used private pilots with limited crew experience performing a less structured task. It may be that flying a 747 generally requires more interaction between crew members than the present task, which would result in better performance being associated with a high level of interaction. In the present study the positive correlation between pilot competitiveness and interaction level may result from a poorer quality of interaction that suppresses the beneficial aspects of interaction, per se. Subsequent analyses of these data may help to clarify the relationships that have emerged in this preliminary report.

References

Foushee, H.C. (1984). Dyads and triads at 35,000 feet: Factors affecting group process and aircrew performance. American Psychologist, 39, 885-893.

Foushee, H.C. & Manos, K.L. (1981). Information transfer within the cockpit: Problems in intracockpit communications. In C. E. Billings & E. S. Chesney (Eds.) Information transfer problems in the aviation system (NASA Report No. TP-1875). Moffet Field, CA: NASA Ames Research Center (NTIS No. N81-31162)

Helmreich, R.L. (1982, August). Pilot selection and training. Paper presented at the annual meeting of the American Psychological Association, Washington, D.C.

Helmreich, R.L. & Spence, J.T. (1978) The Work and Family Orientation Questionnaire: An objective instrument to assess components of achievement motivation and attitudes toward family and career. JSAS Catalog of Selected Documents in Psychology, 8, 35, MS 1677.

Ickes, W. (1981). Sex-role influence in dyadic interaction: A theoretical model. In C. Mayo & N. Henley (Eds.) Gender and nonverbal behavior. New York: Springer-Velag.

Spence,J.T., Helmreich, R.L., & Stapp, J. (1974). The Personal Attributes Questionnaire: A measure of Sex Role Stereotypes and Masculinity-Femininity. JSAS Catalog of Selected Documents in Psychology, 4, 43-44 MS 617.

1. The author's research reported here and preparation of the report were supported by NASA Cooperative Agreement NAG 2-137, Robert L. Helmreich, Principal Investigator.

Mood, Sleep, and Fatigue Effects in Flight Operations

Thomas R. Chidester

The University of Texas at Austin

Abstract

A survey of airline pilots was undertaken to determine normative patterns in mood and sleep during short-haul and transmeridian flights. The results revealed that over the course of a normal multi-day trip, pilots experience a decline in positive mood, or activity, and an increase in negative mood, or tension. On layovers, pilots report getting less sleep and experiencing sleep of poorer quality than at home. These patterns are similar in short-haul and transmeridian operations, but mood and sleep disturbances are more severe among transmeridian pilots. Further, sleep quality and mood are correlated with perceived fatigue and reports of stomach or intestinal distress on trips.

Introduction

Aviation researchers, and those who regulate the airline industry, have been concerned for some time with the impact of demanding flight schedules on pilots. The National Aeronautics and Space Administration (NASA), at Congressional request, is now conducting and sponsoring several studies of the impact of typical flight lines on pilot fatigue and the subsequent effects of fatigue upon performance. A number of studies (reviewed by Chidester, 1985) have found systematic changes and individual differences in fatigue over the course of missions or simulated missions. However, little data exist on baseline or characteristic responses of pilots to their typical schedules. This study presents new data from a large sample of professional airline pilots on changes in sleep quality and quantity and subjective mood over the course of typical flightlines.

The present study complements a large project on fatigue, personality, and performance conducted at NASA-Ames Research Center, and the survey instrument was adapted from NASA's "Background Information Inventory: Studies of Human Performance in Long- and Short-Haul Flight Operations" developed by Foushee and Graeber (1982). The goal of the present study was to produce a self-report survey instrument to obtain perceptions of <u>normative</u> patterns and reactions in order to allow exploration of relationships among sleep, fatigue, and mood while at home and during normal flight operations. Three issues were examined: (1) the extent to which sleep and mood change during flight operations, (2) the relative impact of short-haul and transmeridian operations, and (3) the relationship of these changes with reported fatigue and stomach or intestinal distress.

In this paper, a distinction is drawn between transmeridian and short-haul operations, based upon the number of time zones typically crossed. Short-haul flights are those which involve multiple, relatively short flight segments often crossing 1 but no more than 2 time zones. Transmeridian flights, in contrast, usually involve only one flight segment per day and cross no less than 3 time zones.

Short-haul flights are characterized by a higher consistent work load than transmeridian flights. A crew on a short-haul trip frequently flies in and out of areas of high density operations and on a typical trip, will fly for 3 or 4 days with up to six take-offs and landings per day (Helmreich and Wilhelm, 1984). Transmeridian flights are characterized by long periods of relatively low workload during travel at cruising altitude. On the other hand, air crews in transmeridian operations are subjected to the effects of external desynchronization of biological rhythms caused by the rapid crossing of time zones. When crewmembers layover following rapid transmeridian flight, the social and solar cues by which one regulates one's daily behavior (zeitgebers) are out of synchrony with bodily rhythms (Graeber, 1982). Tepas (1982) reported that the most frequent consequence of this is acute partial sleep disruption and subsequent fatigue.

To examine these issues, a survey research project was undertaken with the cooperation of a major airline. All pilots employed by that airline were asked to complete a survey instrument and their responses were examined.

Method

Respondents. All pilots employed by a major U. S. airline based in the Northeast were surveyed during the month of August, 1985. A total of 763 pilots engaged in short-haul and 110 engaged in transmeridian operations received surveys. Of that number, 463 (60.7%) short-haul and 57 (51.8%) transmeridian pilots responded. The average age of the pilots was 36; the two groups of pilots did not differ in age. This sample is unique in that respect because this airline assigns pilots to type of operations primarily by personal preference. Most carriers assign only high seniority pilots to transmeridian operations, and as a result, short-haul and transmeridian pilots often differ by an average of 10 years of age.

Procedure. The author delivered surveys to the company mail boxes of all pilots. Each survey was accompanied by a cover letter explaining the purposes of the project and a postage-paid return envelope. Pilots were asked to complete the survey and drop it in the U. S. mail. When surveys were received, all information was entered into a secure, deidentified database. After all of the data were entered, the surveys themselves were returned to the respondents.

Survey Instrument. Two different forms of the survey were administered, one for short-haul and one for transmeridian pilots. The forms differed only in the instructional sets given

for each of the mood checklists and in the total number of items. After asking for basic demographic information, the survey requested general information on sleep and fatigue while at home and on overnight layovers, followed by a number of presentations of a 6-item mood adjective checklist. Short-haul pilots were asked to complete the checklist with reference to a typical non-flying day, on the first leg of a flightline following one or more days off, on the evening of an overnight layover, and on the last leg of the second day of flying. Transmeridian pilots were asked to complete the checklist with reference to a typical non-flying day, at the start of a west-bound transcontinental flight, during the evening of an overnight layover, and during the last hour of a returning transcontinental flight. Three additional presentations asked about the beginning, layover, and last hour of a transatlantic round trip (New York-London-New York).

Results

Reduction of Sleep and Mood Data. Two dimensions of sleep were analyzed for home sleeping and overnight layovers. Sleep quantity was straightforward: the number of reported hours of sleep in each setting. Sleep quality, however, was a summed scale of 5 items (alpha=.75). These items were: overall rating of sleep quality, satisfaction with amount of sleep, and frequencies of having difficulty falling asleep, waking-up during sleep, and having difficulty getting-up in the morning. This scale was calculated for both settings. Sleep quality and quantity were moderately correlated (r=.22 at home, .09 on short haul layovers and .24 on transmeridian layovers). Although data on the relationship between sleep and several other dimensions of personality have yet to be analyzed, the impatience factor (Pred and Helmreich, 1985) of the Jenkins Activity Survey (which measures the Type A behavior pattern) correlates -.34 with sleep quality. Two dimensions of mood were calculated for each presentation. The first dimension consisted of the adjectives active, happy, and efficient, and was labeled activity (average alpha=.66). The second dimension consisted of the adjectives annoyed, tense, and tired, and was labeled tension (average alpha=.65). The average correlation between the two dimensions across presentations was -.35 and ranged from -.28 to -.48.

Changes in Sleep and Mood. Pilots engaged in short-haul operations reported sleeping more poorly ($t(463)=9.70$, $p<.001$) and for fewer hours ($t(463)=3.87$, $p<.001$) on overnight layovers than at home. Similar changes were reported by transmeridian pilots (for quality, $t(53)=5.82$, for quantity, $t(53)=4.03$ $p's<.001$). However, effect-size statistics (Glass, 1978) indicated that the changes in sleep quality and quantity were more severe among transmeridian pilots (for quality, $d=0.68$ vs. 0.46; for quantity, $d=0.76$ vs. 0.31). The average amount of sleep lost is about one-half hour.

Among short-haul pilots, activity decreases ($F(3,452)=111.08$, $p<.001$) and tension increases ($F(3,452)=310.09$, $p<.001$) over the course of a flight line. Among transmeridian pilots, similar but magnified trends were reported for transcontinental flights ($F's(3.51)=49.75$ for activity and 26.38

for tension, p's<.001) and transatlantic flights (F's(3,51)=65.85 for activity and 44.69 for tension, p's<.001). Two points can be made in addition to the general trend of increasing tension and decreasing activity over the course of a flightline. First, in each type of flight operation, pilots report higher activity and lower tension when reporting for duty than while at home. That is, pilots feel more excited, happy, or pleasantly aroused when they report for duty than they do at home. Second, and most interestingly, the trends increase linearly in severity from short-haul to transcontinental to transatlantic operations.

<u>Sleep, Perceived Fatigue, and Intestinal Distress</u>. The two dimensions of sleep, while moderately correlated at home and on layover, are differentially correlated with the frequency and severity of reported fatigue and reports of stomach and intestinal distress on flight lines. First, the two dimensions differ in their consistency from home to layover. Sleep quality at home correlated .49 with sleep qualtity on layover among short-haul pilots and .62 among transmeridian pilots. Amount of sleep is less consistent from home to layover (r=.06 among short-haul and r=.22 among transmeridian pilots). Second, amount of sleep is essentially uncorrelated with reported frequency or severity of fatigue (all r's<.10). However, sleep quality strongly predicts frequency (r=-.31) and severity (r=-.14) of fatigue at home and on trips (r=-.34 for frequency and -.29 for severity) among short-haul pilots. Similar but stronger correlations were found among transmeridian pilots.(r=-.46 for frequency and -.38 for severity at home; r=-.49 for frequency and -.52 for severity on trips). Finally, while amount of sleep does not predict frequency of stomach or intestinal distress on trips (all r's<.19 and non-significant), sleep quality on trips correlates -.31 with stomach or intestinal distress among short-haul pilots (the correlation was nonsignificant among transmeridian pilots). This correlation indicates that those who sleep poorly on trips experience gastro-intestinal distress more frequently than those who sleep well.

Discussion

The results reported here lead to three conclusions: First, during layovers following flight operations, pilots experience poorer and less sleep than they get at home. Second, positive mood or activity decreases over the course of a flightline and negative mood or tension increases. Further, pilots are in a better mood when they report for duty than when at home, and the negative changes in mood are more severe in transmeridian than short-haul operations. Third, while both quantity and quality of sleep decrease while on overnight layovers, sleep quality has more important effects on fatigue and stomach or intestinal distress than does quantity.

In interpreting or generalizing these findings, three points should be considered. First, data reported here are normative and subjective, representing aggregations of the experiences of professional pilots. These data are highly reliable, but do not necessarily apply to every trip. Second, the data are as yet unlinked with performance. Performance may or may not parallel

changes in mood or sleep. Foushee (1986), for example, found that crew performance can even improve under certain conditions when pilots are fatigued because pilots recognize the degree of stress they are under and make conscious efforts to work better together. Future research will need to examine the linkages among mood, sleep, and performance.

Finally, a variety of interactions need to be studied, especially those involving time of day. In conversations with the author, pilots who responded to the survey reported that most of the difficulties they experience involve late night short-haul and transmeridian flights. Also, individual differences in personality will probably be found in mood and sleep changes. This report represents only a first cut of the data from the survey and has shed only a little light on a variety of potentially fascinating issues. Further exploration of this dataset should yield a number of interesting relationships between the experience of airline pilots and interactions of personality with aspects of flight operations. The outlook for future research in this area should be very positive.

References

Chidester, T.R. (1985). Mood, sleep, personality, and performance among pilots engaged in short-haul and transmeridian airline operations. Unpublished doctoral dissertation prospectus, The University of Texas at Austin.

Foushee, H.C. (1986). NASA Crew Fatigue Project, personal communication.

Foushee, H.C. and Graeber, R.C. (1982). Background information inventory: studies of human performance in long- and short-haul flight operations. NASA, Ames Research Center.

Glass, G.V. (1978). Integrating findings: The meta-analysis of research. In L. Shulman (Ed.), Review of research in education, 5, 351-379.

Graeber, R.C. (1982). Alterations in performance following rapid transmeridian flight. In F.M. Brown & R.C. Graeber (eds.), Rhythmic aspects of behavior. Hillsdale, N.J.: Earlbaum.

Helmreich, R.L., and Wilhelm, J.A. (1984). Mood, sleep, and personality: A prelimnary report. Paper presented at NASA, Ames Research Center.

Pred, R.S., and Helmreich, R.L. (1985). Construction of a short form of the Jenkins Activity Survey. Unpublished manuscript, The University of Texas at Austin.

Tepas, D.I. (1982). Work/sleep time schedules and performance. In W.B. Webb (Ed.), Biological rhythms, sleep, and performance.. New York: Wiley and Sons.

The research reported here was supported by NASA Cooperative Agreement NAG2-137, Robert Helmreich, Principal Investigator.

FLIGHT DECK MANAGEMENT ATTITUDES: A CLUSTER ANALYSIS

Valerie J. Edwards
University of Texas at Austin

Abstract

A cluster analysis of responses to the Cockpit Management Attitudes questionnaire (Helmreich, 1984) is described. The analysis was performed separately for airline Captains and First Officers. The analyses revealed strong differences between the two groups in their cognitive organization of various issues surrounding flight deck management.

Introduction

Recent analyses of aircraft accidents have led members of the aviation community to focus on the area of cockpit resource management as an avenue to improve aviation safety (e.g., Cooper, White, & Lauber, 1979). Helmreich (1984) described preliminary results of a questionnaire designed to measure beliefs about good cockpit management practices in the multi-crew environment. The items comprising this instrument, the Cockpit Management Attitudes questionnaire (CMAQ), were designed to measure various conceptual and empirical issues and personal reactions related to resource management, and were drawn from line and simulator observations, NTSB accident reports, and NASA research. Initial results from a group of airline pilots revealed that although some consensus on what constituted effective resource management practices existed, substantial response variability was evident in the sample. In particular, an analysis of responses by position yielded highly significant differences between Captains and First Officers, most notably regarding First Officer input.

Further research using the CMAQ has demonstrated relationships between pilots' attitudes and both personality variables and actual performance. Siem (1985) and Helmreich, Siem & Foushee (1985) found relationships between self-reported instrumentality and expressiveness and various attitudes among pilots. Instrumentality here means achievement motivation and self-assertiveness, while expressivity refers to one's interpersonal orientation. Highly instrumental pilots were more likely to endorse items which reflected self-confidence about performing despite situational influences, while high expressiveness was positively associated with the belief that others should question the pilot-in-command's procedures during normal flight.

Helmreich et. al. (1985) reported that responses on the CMAQ reliably discriminated between airline Captains judged "superior" and those rated "below average" by check airmen. Examples of items which differentiated superior pilots include disagreement

with the statement "My decision making ability is as good in emergencies as in routine flying situations", and agreement with the item "Captains should encourage their First Officers to question procedures during normal flight operations and in emergencies".

These studies demonstrate that attitudes toward flight deck management do indeed have a profound effect on performance, implications for the development of training programs in resource management, and, with their demonstrated relationships to stable individual differences, an important role to play in the validation of pilot selection techniques. To these ends, then, reduction of these data into conceptually related subscales should enhance the predictive power of the instrument.

Factor analysis of the instrument using various approaches have yielded equivocal results. Although comparisons between individual questions and measures of personality and performance have yielded interesting findings, defining reliable subscales derived from the 25-item questionnaire is an important refinement for further investigation. Since it was known that strong differences between pilots by position existed, it was hypothesized that these differences may reflect underlying variation in the cognitive organization of these issues into conceptual sets. To test this hypothesis, a cluster analysis of the items from the CMAQ was performed separately for the Captains and First Officers.

Method

From a sample of 698 airline pilots, a subset of all Captains and First Officers were selected for the analyses. The cluster analysis was performed using BMDP cluster analysis of variables (Dixon, 1983). The similarity measure chosen was the absolute value of the correlation between items, and the amalgamation rule for determining similarity employed was single linkage. The analysis proceeds in a stepwise fashion, starting with each variable serving as its own cluster and then combining the two most similar items and continuing to combine items until all variables are joined in one cluster. Single linkage uses the maximum similarity over all pairings of variables as its rule to combine items. This method was chosen since the question of interest was which items each group viewed as most related to each other.

Results

Captains' and First Officers' patterns of clustering were markedly different. The Captains response patterns revealed larger, more distinct and conceptually meaningful clusters, with First Officers having smaller, less well-related clusters. Additionally, the items clustered differed dramatically by position. For Captains, three well-defined clusters emerged. The first cluster deals with cockpit atmosphere, and is composed of the three items dealing with interpersonal sensitivity and

group relations (items 1,2,& 8) plus the two items relating to First Officer input (items 11 and 14). The second distinct cluster within the Captain group deals with issues of leadership (items 6,16,19,21,&10). A third cluster was formed of questions concerning personal invulnerabiity, such as "An effective pilot can leave behind personal problems when flying". A fourth small cluster combines the two items dealing with the positive or negative effects of "chatter" on the flight deck.

First Officers also evidenced a cluster of items concerning flight deck atmosphere, containing the same core of interpersonal items that was included in the Captain's atmosphere cluster, but this cluster did not include the items on First Officer input. Also, First Officers included the item "The pilot flying the aircraft should verbalize his plans for maneuvers and should be sure that the information is understood and acknowledged by the other pilot" in this cluster. Another cluster that was similar to a cluster in the Captain group included the two items concerning cockpit chatter, but also included the item "It is important to avoid negative comments about the procedures and techniques of other crewmembers". Other small clusters emerged such as one combining items concerning First Officer input with a question relating to Captain's management style situations, and an item proposing that the quality of crew coordination is dependent on the Captain's flying proficiency.

Since a stepwise clustering method was employed which results in an eventual recombination of all items into a single cluster, reliabilities were performed on the subscales within each group to determine the appropriate cutoff for the amalgamation process. For the Captain group, Cronbach's alphas of .51 for the atmosphere cluster, .49 for the leadership cluster, and .55 for the personal invulnerability cluster were obtained, using clusters created in five steps or less. For the First Officers, certain items had to be recoded to remove negative relationships so that a reliability analysis could be performed. The Cronbach's alphas for this group's clusters were somewhat lower, ranging from .39 to .48.

Discussion

The analyses described demonstrate that Captains have integrated the concepts and issues associated with cockpit resource management to a greater degree than have First Officers. By empirically deriving subscales through cluster analysis, a clearer picture of the cognitive organization of these attitudes and reactions among airline pilots emerges. The utility of these subscales when applied to the pilot group as a whole (captains, first officers, and flight engineers) is a subject for further investigation, although it seems evident that one would want to apply the subscales derived from the cluster analysis of the Captain group. It seems highly likely that effective crew coordination would be enhanced if all crewmembers shared the same (optimal) cognitive organization of attitudes regarding flight deck management.

REFERENCES

Cooper, G.E., White, M.D., and Lauber, J.K. (Eds.). (1979) Resource management on the flight deck (NASA Report No. CP2120). Moffett Field, CA: NASA-Ames Research Center.

Dixon, W.J. (Ed.)(1983). BMDP Statistical Software. Berkely: University of California Press.

Everitt, B., (1980) Cluster analysis. New York: Halsted Press.

Helmreich, R.L. (1984). Cockpit management attitudes. Human Factors, 26(5), 583-589.

Helmreich, R.L., Foushee, H.C., Benson, B.A., and Russini, R (1985). Cockpit resource management: Exploring the attitude-performance linkage. Paper presented at the 3rd Aviation Psychology Symposium, Columbus, Ohio.

Helmreich, R.L., Siem, F.M., and Foushee, H.C. (1985). Cockpit management attitudes II.: Position, organizational, and personality factors. Unpublished technical paper.

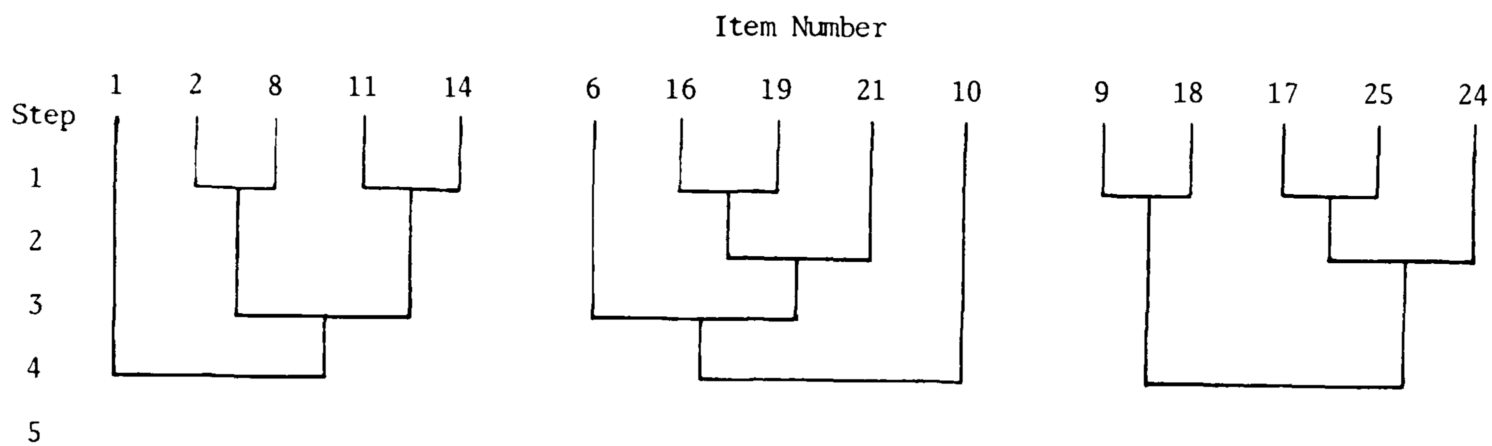
Siem, F.M., and Helmreich, R.L. (1985). Cockpit attitudes and personality in the general aviation community. Paper presented at the 1st Annual Human Factors Society-Alamo Chapter Symposium on Human Factors and Ergonomics, San Antonio, Texas.

The author's research repoted here and the preparation of the report were supported by NASA Cooperative Agreement NAG2-137, Robert L. Helmreich, Principal Investigator.

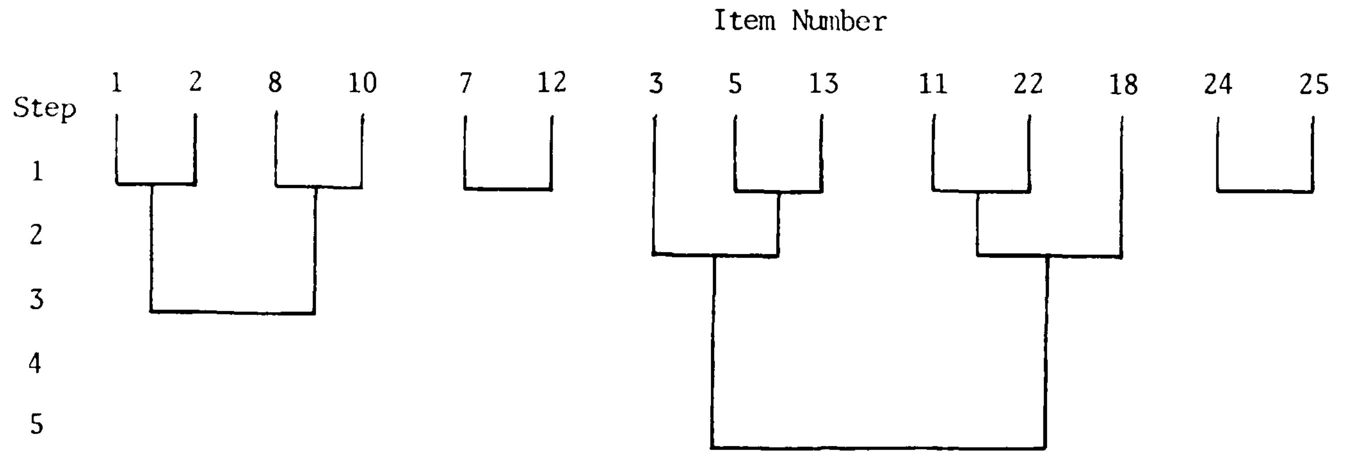
Copies of the Cockpit Management Attitudes Questionnaire may be obtained from Robert L. Helmreich, Professor, University of Texas at Austin, Austin, Texas 78712.

COCKPIT MANAGEMENT ATTITUDES
CLUSTER ANALYSIS
TREE DIAGRAM

Captains



First Officers



Joint-Service Job Performance Measurement/Enlistment Standards Project

D. A. Harris
Office of the Assistant Secretary of Defense
(Force Management and Personnel)

Introduction

The objectives of this Project are to: 1) develop prototype methodologies for the measurement of job performance, and 2) if feasible, link enlistment standards to on-the-job performance.

The individual Services will develop job performance measures and link those measures to enlistment standards within a common methodological framework. This common framework will provide DoD and the Services with technology transfer benefits and considerable cost savings, while preserving Service-specific performance and standards assessment. While many jobs are uniquely defined in a Service, there are other jobs with enough similarity that performance measures developed in one Service can be transferred to another with minor adjustments--and hence, minimal resources expended. Even when performance measures are unique to jobs in a Service, the general methods and techniques for linking those measures to enlistment standards can be quite similar, providing cross-Service applications. This strategy will minimize the time required to develop a comprehensive, DoD-wide program and preclude redundancy in individual Service research efforts.

The primary assumption for this Project is that the technical state-of-the-art in 1) occupational analysis, 2) hands-on performance measurement, and 3) computer technology has advanced to the point that accurate performance measures are feasible and lend themselves to widespread use. In addition, while a major goal of the Project is to provide the necessary data to establish valid linkages between job performance and enlistment standards, another product is accurate, usable, and efficient performance measures which are needed for other manpower, personnel, and training applications. If the performance tests and linkage methodologies are successfully demonstrated on the Project's 28 occupational specialties, in which roughly 20 percent of the force's enlisted personnel are assigned, further research will be undertaken to demonstrate the application and utility of the new technologies for the remaining military occupations.

It is also assumed that the best research strategy is for each Service to conduct its own research in concert with Joint-Service guidelines and timetables. Part of the development effort will be Service-specific and funded with Service research and development (R&D) funds; other efforts will be aimed at Joint-Service demonstrations and funded with Joint Program (6.4) engineering development funds, and the Services' operations and maintenance (O&M) resources. The demonstration projects should be planned to ensure that the data collected will be completed with the ultimate goal of linking performance to enlistment standards, if that goal appears feasible. However, development of specific cross-Service linkage strategies and methodologies are best deferred until the technology to collect the required performance measures has been developed and demonstrated.

The first step of this Project was the selection of military occupational specialties. The need for a single set of common criteria for all Services led to the identification and use of the following seven criteria:

1. The Service-specific military specialties chosen to serve as the test bed for this Project should be of critical importance to the mission of that Service; cross-Service specialties should contain considerable task commonality across the Services.

2. There should be enough people assigned to a selected military specialty to ensure an adequate sample size.

3. The number of individuals within a military specialty assigned to bases should be high enough to provide an adequate sample size without the undue cost of traveling to a large number of bases.

4. Important tasks included in a military specialty should be measurable.

5. Problems within the military specialty (e.g., high attrition) should be known and well documented.

6. Characteristics of the military specialty should permit an evaluation of the impact of the measurement techniques and data collection on minorities and women.

7. The list of military specialties selected should result in a reasonable cross-section of the major aptitude areas measured by the Armed Services Vocational Aptitude Battery (ASVAB) (e.g., electronics, mechanical, administrative, general).

Listed below are the occupational specialties selected by each Service to serve as the test bed for job performance measures development and initial data collection. Collectively, these specialties account for about 20 percent of the total active enlisted force.

<u>Army</u>

Military Police
Administrative Specialist
Cannon Crewman
Motor Transport Operator
Radio Teletype Operator
Vehicle and Generator Mechanic
Tank Crewman
Medical Care Specialist
Infantryman

<u>Navy</u>

Fire Control Technician
Machinist's Mate
Electronics Technician
Electrician's Mate
Aviation Machinist's Mate, Jet Engine Mechanic
Operations Specialist
Radioman

<u>Air Force</u>

Jet Engine Mechanic
Air Traffic Controller
Ground Radio Operator
Electronic Computer

<u>Marine Corps</u>

Automotive Mechanic
Ground Radio Repairman
Infantry Rifleman
Administrative Clerk

Technician
Combat Control Specialist
Avionic Communications
Specialist
Tactical Fighter Repair
Personnel Specialist

Performance Measures Development

The research plan calls for the development of performance measures for an integrated set of cross-Service and Service-specific military occupational specialties that will serve as the initial test bed for demonstrating job performance measurement technologies. Each Service will act as the "lead Service" in demonstrating job performance measures for one cross-Service occupational specialty (e.g., Administrative Clerk). Each Service will also collect a common set of data (e.g., ASVAB scores, training grades, etc.).

Hands-on job performance measures will be developed for all of the military occupational specialties included in the Joint Project. The hands-on test results will then be used as the standard (or benchmark) criterion for the evaluation of substitute (or surrogate) performance measures. The reason for incorporating this feature into the research plan is that hands-on performance tests are both costly and difficult to develop and administer. The Services will not be able to support operationally the development and use of hands-on, benchmark performance tests for all remaining military occupational specialties. Therefore, the Services will develop surrogate measures to see how well they compare with the hands-on, benchmark measures. In this way, the Services will be able to determine the surrogate's utility as a less expensive and less difficult to administer measure of job performance.

The Army, Navy, and Air Force will be responsible for developing specialized expertise in the following surrogate performance measures:

Army

1. Army-wide performance measures. Army-wide performance measures are intended to be indicators of general performance and effectiveness not related directly to the performance of jobs or job-specific tasks. They include such indices as supervisory and peer ratings of soldier effectiveness.

2. Job knowledge, paper-and-pencil tests. These paper-and-pencil tests measure an individual's job knowledge of the critical steps in representative tasks of a specific occupational specialty.

Navy

1. Simulation and training device performance tests. In those Navy occupational specialties (ratings) for which training simulators are available, job performance tests will be developed which take advantage of the simulator's unique, high-fidelity measurement capabilities to assess the performance of Navy personnel on critical tasks of specific ratings.

2. Symbolic simulation tests. These substitute tests are videodisc and/or paper-and-pencil instruments (e.g., two dimensional pictures of equipment)

that link performance test items to the critical steps in representative tasks of the job.

Air Force

1. Walk through performance tests. Walk through performance tests combine hands-on measures with interview testing. Interview testing is a task-level job performance measurement system which uses structured questionnaires scored by trained interviewers to evaluate oral performance descriptions made by job incumbents of critical tasks which do not lend themselves to hands-on testing.

2. Task-level ratings of performance. These ratings reflect the evaluations of supervisors, peers, and incumbents on how well individuals perform on the job. These ratings include evaluations of task-specific performance on the job as well as more global measures of technical and social/interpersonal competence on the job.

3. Task-level ratings of job experience. These ratings provide a measure of the relative amount of experience an incumbent possesses on specific tasks of the job.

Marine Corps

The Marine Corps will be responsible for identifying peripheral data (e.g., biographical, demographic and training data) and monitoring their collection.

Data Collection and Analyses

Once performance measures have been developed, the Services will conduct field tests to demonstrate the measures' accuracy, validity and reliability. To ensure that the field tests will provide the Services with useful data, the following data collection and analysis strategies have been established:

1. To the maximum extent possible, data collection will be conducted in a manner that will allow research results to be used by the other Services.

2. Appropriate Service-specific data banks will be established within each Service to support job performance data collection and analysis efforts.

3. Within guidelines established by the Services, summary cross-Service job performance information will be shared and stored centrally within the Department of Defense.

4. Comparative studies will be conducted by the Services to determine which performance measures and linkage methodologies yield the most accurate, valid and reliable information.

5. Cost/benefit analyses will be conducted to determine which of the surrogate measures produce the most cost-effective information and results.

Conclusion

Determining manpower quality requirements with any measure of precision is not currently possible. Such an effort would require a validated cost trade-off model that would permit the evaluation of the relative costs and benefits of different force quality levels. We know that high quality personnel are more costly to recruit yet less costly with regards to training, force size or effectiveness. A trade-off analysis could vary the quality mix and assess the total costs for recruiting, training, and force maintenance holding constant some performance objective. The question would then become whether there is some optimal quality mix that minimizes the total cost of recruiting, training, and force maintenance for a given outcome. While such a model is theoretically possible, it is still in the developmental stage and is not yet an appropriate basis for policy decisions.

A principal problem in developing a validated cost trade-off model is the difficulty in measuring on-the-job performance. The basic instrument used by the Services to select enlisted personnel is a paper-and-pencil test called the ASVAB. Traditionally, the ASVAB has been validated against training success. The Joint-Service Project is developing prototype job performance tests in selected jobs in each of the Services; cumulatively, most types of military specialties will be represented. Once reliable job performance measures are available, the Joint-Service Project will examine the relationship of job performance measures to the ASVAB. The ultimate goal will be to validate quality enlistment standards against actual job performance, instead of success in training. If this research enables us to establish a definitive link between enlistment criteria and job performance, we will then be able to determine accession quality requirements with greater precision than is now possible.

The Military Services are working together on this problem and the preliminary results of the Project are promising. Demonstration of prototype performance measures and initial attempts to link those measures with enlistment standards are expected in FY 1987. Assuming success, the Project will begin to institutionalize the results of this demonstration in the enlistment process for the specialties selected in this Project and will explore the usefulness of performance data in the training, personnel management, and occupational analysis areas. To do this, the Project will address such issues as the generalizability of performance data. If successful, the Project will ensure the maturation of Project data beyond the enlistment standards application and enhance the long-term use of such data in the new areas. The Project will then begin to demonstrate the application and utility of the new technologies for other occupational specialities.

ARMY RESEARCH TO LINK STANDARDS FOR ENLISTMENT TO ON-THE-JOB PERFORMANCE

Jane M. Arabian and Lawrence M. Hanser
U.S. Army Reserch Institute for the
Behavioral and Social Sciences

Overall Army Goals

The Army Research Institute is currently engaged in a large-scale, multiyear project to improve the Army selection and classification system and, thereby, increase the overall effectiveness of the force. This research is being conducted with contractor support from Human Resources Research Organization, American Institutes for Research, and Personnel Decisons Research Institute. The goal of the Army's program for increasing the efficiency of enlisted personnel selection and utilization is to enable the Army to meet its peacetime and mobilization missions through improved matching of individuals to military occupational specialties (MOS). The research is aimed at developing comprehensive selection and classification procedures to validly predict performance in Army training and occupational specialties. Specifically, this project will:

1. Validate existing selection measures against both existing and project-developed criteria. The criteria will include both Army-wide performance measures based on newly developed rating scales and direct measures of MOS-specific task performance.

2. Develop and validate new and/or improved selection and classification measures.

3. Validate intermediate criteria, such as performance in training, as predictors of subsequent criteria, e.g., job performance ratings, so that informed reassignment and promotion decisions can be made throughout an individual's tour.

4. Determine the relative utility to the Army of different performance levels across MOS.

5. Estimate the relative effectiveness of alternative selection and classification procedures in terms of their validity and utility for making operational selection and classification decisions.

Military Occupational Specialties and Performance Measures

The Army's research focuses on 19 MOS. The MOS were selected to be representative of the Army and include all operational ASVAB aptitude area composites. Blacks, whites, Hispanics, males, and females are present in these MOS in the same proportions as in total accessions. These MOS represent 44% of annual Army enlistments.

A number of performance measures, including measures of training success, service-wide performance, and MOS-specific hands-on performance, will be developed for these MOS. For reasons of cost efficiency, not all measures will be developed for the 19 MOS. All project criterion measures will be developed for the following MOS: Cannon Crewman (13B), Motor Transport Operator (64C), Administrative Specialist (71L), Military Police (95B), Infantryman (11B), Tank Crewman (19E), Radio Teletype Operator (31C), Light Wheel Vehicle/Power Generation Mechanic (63B), Medical Specialist (91A).

Measures of training success and service-wide performance will be developed for the following specialties: Supply Specialist (76Y), Food Service Specialist (94B), Combat Engineer (12B), MANPADS Crewman (16S), Ammunition Specialist (55B), Petroleum Supply Specialist (76W), Chemical Operations Specialist (54E), Utility Helicopter Repairman (67N), Carpentry/Masonry Specialist (51B), TOW/Dragon Repairman (27E).

The Army's rationale for the development of multiple measures of job performance is based upon the knowledge that a soldier's job is multi-faceted and there are multiple aspects to job performance. Therefore, the Army's research project has developed different kinds of tests to assess these different aspects of job performance and thereby obtain information about the domain of job performance behaviors.

A complete report of accomplishments and current status of the Army's research to link enlistment standards to job performance has been published by the Army Research Institute (1983, 1984, 1985) and is available for distribution. What follows is a summary of accomplishments in the area of job performance measurement during CY85.

Performance Measures Development

Job knowledge tests, hands-on tests, and job performance and Army-wide rating instruments were developed in CY84 for MOS 13B, 64C, 71L, and 95B. A report was prepared describing the rationale for and procedures followed in the analysis and selection of relevant job tasks for the hands-on and job-specific knowledge tests for these MOS (HumRRO and ARI, 1984; Campbell and Harris, 1985). Following the recommendations in that report, tests and rating instruments were developed in CY85 for the remaining MOS.

Task Selection and Instrument Construction. A detailed, technical account of the procedures for task selection and instrument construction is documented in Campbell and Harris (1985) and the FY85 Project A Annual Report (1985). A brief account is presented below.

The general model and procedures for performance measure criterion development in Project A is as follows: The basic cycle of a comprehensive literature review, conceptual development, scale construction, pilot testing, scale revision, field testing and proponent (management) review was followed for each kind of criterion measure. The primary goals of criterion measurement in Project A were to: a) make a state-of-the-art attempt to develop job sample or "hands-on" measures of job task proficiency, b) compare hands-on measurement to paper-and-pencil tests and rating measures of proficiency on the same tasks (i.e. a multi-trait, multi-method approach), c) develop rating scale measures

of performance factors that are common to all first tour enlisted MOS (Army-wide measures), d) develop standardized measures of training achievement for the purpose of determining the relationship between training performance and job performance, and c) evaluate existing archival and administrative records as possible indicators of job performance.

For the hands-on measures, a comprehensive task sampling procedure was used to define the population of tasks in each MOS and then select 30 job tasks to represent the population of the MOS tasks. The task lists were then reviewed by the proponent schools for completeness and representativeness of the occupation. Fifteen tasks requiring a high level of physical skill, a series of prescribed steps, and speed of performance were selected for hands-on testing. The test items for the hands-on measures were generated from training manuals, field manuals, interviews with officers and job incumbents, and other appropriate sources.

The job knowledge tests, a paper and pencil multiple choice format, were developed to cover all of the thirty tasks in the MOS lists. The item content was generated on the basis of training materials, job analysis information, and interviews.

Two types of rating scales were also developed. One type of seven-point scale was designed to be parallel to the job tasks that were measured in the hands-on mode; one scale was developed for each of the fifteen tasks. The second type of rating scale followed standard procedures for developing Behaviorally Anchored Rating Scales from "critical incident" workshops involving 70-75 officers and NCO's. This procedure resulted in six to nine MOS-specific Behaviorally Anchored Rating Scales, or BARS, for each of the nine MOS. A similar procedure was used also to develop Army-wide performance rating scales.

Training Knowledge (achievement) tests were also developed. These tests were based upon training course content. The content distribution of items on the test was proportional to the content of the course. The item pool was written by a team of subject matter experts, contracted for that purpose; the items were edited for clarity and relevance to training and job performance prior to field testing.

Test Administration. Each test site had a test site manager who supervised all of the research activity and maintained the orderly flow of personnel through the data collection stations. An officer and two NCOs from one of the supporting units at the test site was assigned to support the field test. The officer provided liaison between the data collection team and the tested units; the NCOs coordinated the flow of equipment and personnel through the data collection procedures.

During the week preceeding data collection at each research site the NCO scorers for the hands-on measure were given two days of training on scoring procedures by members of the research staff. The scorers were told about the overall design and nature of Project A; their critical influence on the reliability and validity of the measures was emphasized. Test administrators (contractor research staff and Army civilian research staff) for the remaining criterion measures were also trained at that time on the test administration procedures and one test administrator was assigned to each test station.

For data collection purposes, the criterion measures were divided into four major blocks:

1) Hands-on measures
2) Rating measures
3) Paper and pencil job knowledge tests
4) Paper and pencil training achievement tests.

Each block comprised one half day of participant time and each participant was tested for a two-day period. The order of administration of the blocks of measures was counterbalanced at each test site. Data collection at each site required approximately two weeks to complete.

Lessons learned from the field testing fall into three broad categories: logistics, scorer and test administrator training, and test site management. Based on the experience with the field tests, procedural modifications were implemented for the concurrent validation. Modifications include, but are not limited to: earlier coordination with Army support personnel for the test site; use of the Army-wide personnel locator to name-request soldiers for testing, rather than designating units to supply all their appropriate soldiers; intensive one-week training course for test administrators prior to arrival at test sites; development of a hands-on scorer certification program; and improved systems for conducting the orderly flow of personnel through data collection stations.

Reliability Results. Early reliability analyses of the field test data were used to provide information for the subsequent revision of the criterion measures to be used in the concurrent validation. Accordingly, measures were revised, if needed, to improve their statistical reliabilities and factor structures. This revision process included the elimination or modification of some test items; however, the revisions did not compromise the range of item difficulty (easy to hard) present in the measures. Following the revision and prior to the concurrent validation data collection, each measure was officially reviewed for content and accuracy and approved by the Commanding General with proponency for the respective MOS.

The results of the statistical analyses are presented in detail in the field test reports (1985) and the FY85 Project A Annual Report (1985).

Performance Measures and ASVAB Relationships. Analyses of the interrelationships among the different types of performance data (from hands-on, written knowledge, and rating measures) resulted in moderate correlations. The results indicated that different aspects of the job performance domain were being measured by the different testing instruments, as intended. Very high correlations would have indicated that the instruments were all measuring the same aspect of the performance domain. Very low correlations, on the other hand, might have indicated that the data were primarily a reflection of the different measurement methods. There were no plans to examine the relationships among the performance measures and ASVAB during this stage of the project. The relationships between the performance measures and ASVAB, as well as continuing analyses among the performance measures, will be examined closely using the data collected in the concurrent validation phase of the research project.

REFERENCES

Campbell, J.P., and Harris, J.H. (1985, August). Criterion reduction and combination via a participative decision making panel. Paper presented at the Annual Convention of the American Psychological Association, Los Angeles, CA.

Human Resources Research Organization (HumRRO) and American Institutes for Research (AIR). (1984). Selecting job tasks for criterion tests of MOS proficiency. Research Report in press; Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Human Resources Research Organization (HumRRO) and American Institutes for Research (AIR), Personnel Decisions Research Institute (PDRI), and Army Research Institute (ARI). (1983). Improving the Selection, Classification and Utilization of Army Enlisted Personnel: Annual Report. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Human Resources Research Organization (HumRRO) and American Institutes for Research (AIR), Personnel Decisions Research Institute (PDRI), and Army Research Institute (ARI). (1984). Improving the Selection, Classification and Utilization of Army Enlisted Personnel: Annual Report. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Human Resources Research Organization (HumRRO) and American Institutes for Research (AIR), Personnel Decisions Research Institute (PDRI), and Army Research Institute (ARI). (1985, in preparation). Improving the Selection, Classification and Utilization of Army Enlisted Personnel: Annual Report. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Human Resources Research Organization (HumRRO) and American Institutes for Research (AIR), Personnel Decisions Research Institute (PDRI) and Army Research Institute (ARI). (1985, in preparation). Report on the Results of the Field Tests. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

NAVY JOB PERFORMANCE MEASUREMENT PROGRAM

Gerald J. Laabs, Herbert G. Baker, Leonard Kroeker, and Pamela Kidder
Navy Personnel Research and Development Center
San Diego, California 92151-6800

Abstract

As part of the Joint-Service Enlistment Standards/Job Performance Measurment Project, the Navy is developing job performance measures and demonstrating their use as criteria for predictor validation. Hands-on and substitute performance measures will be developed for six ratings including Machinist's Mate, Radioman, Electronics Technician, Operations Specialist, Fire Control Technician, and Electrician's Mate. The substitute measures will consist of a low fidelity computer based simulation or a paper and pencil simulation, and behaviorally anchored rating scales. Measures will be administered to approximately 300 first-term enlistees. Data will be analyzed to determine the reliability and validity of the job performance measures.

The Navy's contribution to the Joint-Service Enlistment Standards/Job Performance Measurement Project is entitled Performance-Based Personnel Classification. The main objective of our project is to investigate measurement approaches that might be used to make the Navy's automated classification and assignment system more performance-based.

Our primary goal is to develop measures of job performance and demonstrate the use of these measures as criteria for predictor validation. In accordance with the Joint-Service Project research strategy, we are constructing measures of technical proficiency and relating these measures to selection instruments.

Our secondary goal is to lay the foundation for further research to improve the capability of the Navy in measuring human performance. Measurement capabilities are being developed in the areas of aptitude and job performance, including (1) the development of a job performance data base, (2) the investigation of new predictors, and (3) the transfer of measurement technology across Services.

Research Plan

The focus of the Navy research plan is on: (1) establishing relationships among three types of measures of technical proficiency, that is, a job sample test, a job sample simulation, and a set of rating scales, and (2) evaluating the predictive relationships between these measures and subtest scores on the Armed Services Vocational Aptitude Battery (ASVAB).

The work has been partitioned into five tasks:

1. Develop a prototype rating-specific data base that contains job performance measurement and prediction information.

2. Identify candidate ratings for which job performance measures will be developed.

3. Design, develop, and field job performance tests and instruments.

4. Investigate the feasibility of a performance-based factor or component for classification and assignment.

5. Demonstrate the potential usefulness of new predictors being developed in related technology-base research.

Data Base Development. This work serves two purposes: one is to ensure that we proceed as economically as possible in this project; the other, to prepare for future research regarding ability and performance measures. Existing assessment programs and task analysis data bases have been reviewed, and training simulators have been evaluated as sources of job performance information. This information will be the basis of a prototype rating-specific data base system that can be used to support the measurement and prediction of job-related performance. Because of the substantial interest shown in this task, it has been significantly increased in scope and has become a separate project.

Ratings Selected for Study. Six candidates were selected from two lists of Navy ratings that were rank ordered on the basis of (1) the criticality of each rating with respect to the accomplishment of the Navy's missions, and (2) the number of individuals assigned to each rating. These ratings also use five different ASVAB selector composites for assignment into the job:

1. Machinist's Mate Rating (MM)

2. Radioman Rating (RM)

3. Electronics Technician (ET)

4. Operations Specialist (OS)

5. Fire Control Technician (Surface) (FC)

6. Electrician's Mate (EM)

The following seventh rating was selected for a special Joint-Service Project demonstration in which Air Force job sample tests and substitutes are being applied to the corresponding Navy job:

7. Aviation Machinist's Mate, Jet Engine Mechanic (AD).

The Joint-Service research strategy will be applied to these seven ratings -- each of which is really a project in itself. It is anticipated that after the seven projects are completed, we will have sufficient evidence to carry on work using one of the substitute measures, thereby eliminating the need for more testing using the expensive hands-on type of measure.

Types of Measures. To achieve high fidelity in a proficiency measure, we have to observe whether an individual "can do" a job. The closest that we are able to come to observing on-the-job performance, is to develop a hands-on test that samples important job tasks and uses actual equipment in the testing. Because such tests are extremely expensive to construct and administer, the Navy is investigating two types of substitutes: (1) simulations of job samples, and (2) rating scales.

One type of substitute for the hands-on test will use either "low" fidelity computer-based simulation technologies or paper-and-pencil simulation techniques that rely on pictures and illustrations of actual equipment. The second type of substitute will use behaviorally anchored rating scales that specify work behaviors representing different levels of proficiency. In addition, a set of behaviorally anchored rating scales will be developed at the more general job performance level.

A hands-on performance test covering critical job tasks will be developed for each rating to be studied. Field-test packages that include the hands-on test and the two types of substitute measures have been developed for the first two jobs, that is, the MM and RM ratings. Test development is currently under way in the ET rating.

New Predictors. The purpose of this work is to examine new measures of ability adapted from other projects, in order to demonstrate the usefulness of new predictors for personnel selection and classification.

Although, it was originally planned to conduct this work late in the life of our project, it is progressing at a very fast rate. An increasing amount of interest is being shown in the development of new predictors to supplement ASVAB, especially with the advent of the Computerized Adaptive Testing (CAT) system. Therefore, this task has been significantly increased in scope and has become a separate Joint-Service project.

Data Collection

A field-test package including all three types of measures described above will be administered to a sample of approximately 300 first-term enlistees from the fleet. For most ratings, the field-test package will be administered on board ship and in pierside buildings or vans. Testing is underway in the MM rating and has just been completed in the technology transfer study, which involved members of the the AD rating and their Marine Corps counterparts.

The total time required to test one examinee is running about 12 hours. This total time is divided among: (1) A hands-on, job sample test taking from four to six hours, (2) a job sample simulation taking from one to three hours,

(3) a new predictor test taking about two hours, and (4) sets of rating scales taking about one hour. The simulation and self ratings are administered at the beginning of the testing period. A set of rating scales is also administered to at least two peers and one first-line supervisor of each examinee.

Additional information is collected about each examinee such as time on job task and personnel record entries. Periodically, a Navy representative acts as a second observer to ensure that contract testing continues in accordance with training and to reinforce proper data collection procedures.

Data Analysis

The data collected using the field-test package will be analyzed to see which of the alternative measures can serve as a substitute for the hands-on test. The performance data will ultimately be related to new and existing selection instruments. Test reliability and validity will be estimated using a variety of analytic procedures.

Reliability. The use of hands-on performance tests raises unique problems in evaluating test reliability. Since such tests are scored by observers, reliability depends on the objectivity of their observations. To identify and reduce observer subjectivity, the Navy standardized scoring procedures and trained observers for the first ratings currently being tested.

Test objectivity was determined by estimating the degree of agreement among observers across examinees and tasks. We found extremely high agreement for the MM and AD rating tests: Correlations were all above the .95 level and many times there was complete agreement between the paired observers. New analytic procedures will also be used by the Navy to examine the data from paired observers. For example, these data are being analyzed using generalizability theory designs to estimate the variance associated with observers and other facets that may affect performance scores.

Validity. The validities of the hands-on performance test and its substitutes depend on the representativeness of the tasks included. The adequacy and completeness of the job content representation or content validity derives from the task selection procedures. In addition to the usual selection procedure based on subject matter expert judgments of such job task charateristics as criticality, the Navy has also studied probability sampling to arrive at a true job sample from which performance by a limited number of members of a rating can be generalized to the entire rating.

To further evaluate the tests and instruments, the following analyses will be conducted on the data collected from the entire sample in each rating: (1) Examination of the distribution of scores, (2) item analyses, (3) interobserver reliability using the data from a Navy representative as the second observer, (4) interrater reliability for peer ratings, (5) interrater reliability across self, supervisor, and peer rating sources, and (6) multitrait-multimethod analyses for multiobserver data, ratings from different sources, and the three different measurement methods. A final series of analyses will use generalizability theory designs.

Improving Personnel Classification and Assignment

The Navy currently uses an automated classification and assignment system to place recruits into ratings or jobs. Ultimately, the relationships uncovered between selection instruments and measures of job performance will be used to make this system more performance based.

The Navy's automated classification and assignment system contains several algorithms designed to achieve the best person-job match within the constraints of fluctuating recruit quality and varying occupational requirements. The system is called CLASP, which stands for Classification and Assignment within PRIDE (which is the Navy's computerized school-seat reservation network).

The CLASP system is a real time conversational computer program that provides classifiers with a list of ratings appropriate for the recruit being processed. Data describing a recruit are entered entered into the system and an optimality index value reflecting the match between the recruit and all ratings is computed. A subset of ratings with the highest values is then produced. Having the recruit and classifier concentrate on this list of ratings, ensures that recruits are assigned in a near optimum manner.

At the present time, the optimality index values of CLASP are based on the following factors:

1. Prediction of school success

2. Match between aptitude and job complexity

3. Match between preferences and manning priorities

4. School quotas

5. Minority representation

6. Prediction of attrition

To make personnel classification more performance-based, one or more factors that forecast job performance must be added to the above list. Before we can extend job performance measurement technology to the total Navy enlisted force in support of personnel classification, however, we have to demonstrate successful linkage to enlistment standards for the subset of ratings selected for study. This must include the demonstration of reliable and valid substitutes for hands-on performance measures, because the use of hands-on performance testing to gather criterion data on a routine basis is prohibitively expensive and time consuming. In addition, significant and practical relationships between ASVAB scores and job performance data must be demonstrated. If we are successful, improved person-job matching will result.

Air Force Job Performance Measurement Research

Rodger D. Ballentine
Air Force Human Resources Laboratory

Abstract

The Air Force is participating in a congressionally mandated Joint-Service Project to develop a technology for measuring job performance and determining the feasibility of linking enlistment standards to job success. This paper describes the background, research design, and status of the Air Force criterion development effort, as well as plans to institutionalize this technology.

Background

The Air Force Human Resources Laboratory (AFHRL) is developing a technology for systematically obtaining job performance measures. The short-term objective of this effort is the development of on-the-job performance measures to validate Air Force selection and classification procedures for first-term airmen. Guidelines for developing and obtaining performance measures for officer and civilian jobs will also be established. Once obtained, these measures will be placed in a data base for validation use. The long-term goal is to establish an operational performance measurement program so the measures will be available for evaluation of selection and training procedures and personnel policies and practices.

Planning for the Air Force's program of research in performance assessment began several years ago as the result of three primary requirements. Military and civilian manpower, personnel, and training managers asked the AFHRL to develop an approach for measuring job performance so that the measures could be used to assist in the evaluation of their training and selection programs. Secondly, the research community needed performance measures to serve as criteria in their research projects. Plans for the Air Force performance measurement effort to meet these requirements were already under development when a third requirement for these measures came with the congressional mandate to test the feasibility of validating the Armed Services Vocational Aptitude Battery (ASVAB) against job performance measures.

The Air Force research plan has been developed in coordination with the other Services to ensure that all programs are interlinked and not redundant. An integrated set of cross-Service and Service-specific enlisted specialties have been designated to serve as the initial test bed for the development of job performance measurement strategies. The following eight Air Force Specialties (AFSs) were selected for the Air Force component of the Joint-Service Project: Jet Engine Mechanic (426X2), Air Traffic Control Operator (272X0), Avionic Communications Specialist (328X0), Ground Radio Operator (293X3), Electronic Computer Specialist (305X4), Combat amd Control Specialist (274X0), Tactical Aircraft Maintenance Specialist (431X1), and

Personnel Specialist (732X0). These AFSs cover the range of jobs in the mechanical, electronic, general, and administrative aptitude areas. Further developmental efforts are planned for eight more enlisted AFSs from each aptitude area, as well as selected officer and civilian specialties. In addition, because each Service has chosen a unique performance measurement orientation, technologies developed by the Air Force in cross-Service specialties can be transferred to other Services.

The overall program of research calls for the development of measurement techniques that will allow for the collection of valid, accurate, and reliable hands-on job performance information. These measures in turn will be used as benchmarks against which surrogate indices of performance (less expensive, easier to administer interview tests and performance ratings) will be evaluated as substitutes for the more expensive, labor intensive hands-on performance measures. In addition to newly developed performance measures, existing measures such as technical training scores, Airman Performance Report ratings, and skill-level advancement indices will also be evaluated as possible surrogate measures.

The rationale for this multi-faceted approach to performance measurement involves cost considerations (the hands-on measures are very expensive to develop and administer), but it is also based on the hypothesis that the various methods, while to some degree overlapping, also measure different aspects of job performance with differing levels of accuracy. Measures of job experience and motivation will also be developed to assess their affects on job proficiency.

Types of Measures to be Developed

Walk-Through Performance Testing (WTPT) is a task-level job performance measurement system that expands the range of job tasks an individual is measured on by combining hands-on task performance and interview procedures to provide a high fidelity measure of an individual's technical job competence. The hands-on component resembles a traditional work sample designed to measure performance on a sample of tasks that have survived the imposition of essential measurement constraints such as testing time/cost or risk of personal injury/equipment damage. The interview testing component has been added as a means of assessing those tasks that would have been eliminated because of these constraints to provide a more thorough coverage of the work domain.

Interview testing takes place in the work setting and requires the evaluator to assess proficiency on a task by asking the incumbent to "show-and-tell" how the task is accomplished. This procedure is designed to uncover procedural strengths and weaknesses and knowledge related to the performance of that task. There will be an overlap between tasks measured by interview and hands-on to determine if the two methods provide similar results. The interview testing component will be evaluated both as a supplement to hands-on measures and as a more cost-effective surrogate.

A wide range of rating forms will be developed as alternative job performance measures to assess both task-oriented and general competence of first-term airmen. These include peer, supervisor, and self ratings at four different levels of measurement specificity: task, dimension, global, and Air

Force-wide. A final set of data collection instruments assesses job experience and level of motivation so these effects on job performance can be measured and then taken into account while studying the accuracy of the performance measures and the relationship between selection tests and job performance.

Construction Strategy

Development of WTPT and AFS-specific rating forms is based on information available in the Air Force Occupational Research Data Base. This data base contains current task-level information on all major enlisted specialties collected through the Air Force Occupational Survey (Job Analysis) Program. Initial work begins with WTPT because it is the most detailed measurement technique, followed by rating form development.

Occupational survey data are utilized in the selection of WTPT tasks. Once critical tasks have been selected, subject matter experts (SMEs) aid test developers in dichotomizing the tasks into those which can be measured by hands-on and those which must be measured by interview. Next, test developers and SMEs outline procedures for observing task performance, interviewing, and specifying the performance standards for scoring responses. Another group of SMEs reviews the instruments developed to assess work performed in a specialty and administration and scoring procedures.

Using information learned in the WTPT development process, rating forms varying in detail from the task-level to global are developed. The most detailed rating forms have a one-to-one correspondence with the tasks in the WTPT. Rating forms are also developed to reflect groupings of tasks, or dimensions, of technical proficiency. In addition, a global rating form provides ratings of technical and interpersonal competences. Finally, an Air Force-wide scale allows evaluation of general competencies required of individuals throughout AFSs.

Questionnaires are also designed to assess job experience and level of motivation. Job experience questionnaires are designed to collect task-level experience on all tasks in the performance test as well as general work experience of the incumbent. Finally, a motivation questionnaire is developed to assess the level of motivation the individual brings to the testing situation.

Once the candidate measures have been developed, they are pretested prior to full-scale data collection. The assessment package is administered in the field to a small sample of job incumbents, their peers, and supervisors. Results are evaluated in terms of testing time estimates, procedural stability, and tester/rater reliability. This scenario is repeated in each specialty during development of the performance measures.

Data Collection and Analyses

Whenever possible, at least 300 first-term airmen who have been on the job at least three months are sampled based on selected demographic characteristics. In some AFSs, the sample is limited further to airmen working in certain types of jobs because the work domain is broad and heterogeneous, encompassing many jobs. In such AFSs, measurement development

and sample selection are limited to those jobs most representative of the specialty (i.e., those jobs in which the majority of the incumbents work).

In all, there are four sources of job performance measures: WTPT examiners, first-line supervisors, peers, and the job incumbents. Examiners observe and evaluate performance using the WTPT methodology, while the raters evaluate an individual's proficiency on the basis of recent past performance. The WTPT examiners, if possible, are former enlisted Air Force personnel with previous job experience in the AFS being studied who have no prior knowledge of the examinee's performance. Extensive WTPT training is provided to examiners. Experienced SMEs are employed in the development of both the WTPT procedures and training. Supervisors, peers, and job incumbents receive training to familiarize them with the purpose and use of rating forms. This training is provided just prior to administration of the rating forms.

To the maximum extent possible, WTPT takes place at the work site where the tasks are actually performed, or at installation training facilities where required equipment is available. Such real-world settings enhance confidence in the adequacy of the testing procedures and performance evaluations obtained. Supervisor, peer, and self ratings are collected by group administration in large centralized locations. The characteristics of testing locations are as consistent as possible across test sites.

Early analyses will focus on indices of inter-rater reliability and psychometric effects (i.e., halo, leniency, range restriction). A multiple linear regression approach will identify relevant variables to be taken into account or controlled and will identify performance measure relationships with ASVAB scores. Where normative performance standards exist (e.g., WTPT), accuracy indices will also be calculated for the surrogate measures.

In the early years of the project, the goal is to develop and compare measurement methodologies. It is probable that some AFSs may require entirely different combinations of measures. Also, some measurement techniques which serve as benchmarks will be prohibitively expensive if used on a large scale. Therefore, speciality-specific and across-specialty trend analyses will be conducted to identify the most cost-effective combinations of measures.

Project Status and Accomplishments

Data collection has been completed on one specialty (AFS 426X2) and instruments pretested on three more specialties (AFSs 272X0, 328X0, and 293X3). The data collection for these next specialties will begin in CY86, as will development of the instruments for the final four specialties (AFSs 305X4, 274X0, 431X1, and 732X0). In addition to the work in these specialties, the Air Force completed the first transfer of technology to other Services by adapting jet engine mechanic instruments for the J-79 engine to the Navy and Marine Corps. In CY86, the Navy will administer these instruments to Navy and Marine jet engine mechanics to determine if similar relationships occur between surrogates and hands-on measures across Services.

Additional research on the validity and accuracy of performance measures has been completed with several new studies initiated to improve the training of evaluators and accuracy of performance information. The Air Force also conducted a preliminary study to examine how performance measurement

information could be used to enhance training effectiveness. In addition, a special evaluator study was conducted to determine if differences exist between WTPT ratings rendered by contractor evaluators, active duty evaluators from the location of the testing, and active duty evaluators from a location other then where the testing is being conducted. The results of this evaluator study could have a substantial impact on data collection for future specialties, including operational plans.

Institutionalizing Job Performance Measurement Technology

The Air Force's current research plan for the enlisted force is to develop performance measures for 40 specialties (8 for the Joint-Service Project) covering a range of mechanical, electronic, general, and administrative specialties (10 per aptitude area). Hands-on performance measures will be developed for the 8 Joint-Service specialties (2 in each aptitude area). It is presumed that the additional 32 specialties would be measured using appropriate surrogate assessment techniques. The surrogate techniques selected will be determined in FY89 using the results of the Joint-Service demonstration project. Measurement of these 40 specialties will provide the Air Force a performance data base for approximately 16% of the specialties which constitute almost 40% of the first-term airmen.

If the technology is successful, cost-effective job performance measures could be obtained for the remaining Air Force specialties through an operational program beginning in FY92. In such a program, surrogates (i.e., rating forms) would be the least costly option with hands-on performance measures of job family clusters being the highest cost option. The Air Force does not plan to develop hands-on performance measures on all specialties due to excessive costs and belief that efficient combinations of surrogate tests can be identified which meet measurement accuracy and cost constraints.

Applications of Technology

If Air Force use of job performance measures in validating selection procedures and setting enlistment standards is feasible, a specialty specific performance measurement data base could be kept current and used routinely to validate those selection tests and procedures. Regardless of the outcome of the standards setting feasibility issue, the performance measurement technology and resulting data could be used by the personnel and training research and development (R&D) community to evaluate their research products, the operational manpower and personnel community to evaluate the impact of personnel policies and procedures, and the operational training community to develop/evaluate training.

The Air Force's most promising application of this technology outside of the enlistment standards project is in training development/evaluation. The training interface has been specifically requested by HQ Air Training Command and initial investigation of the applicability of this technology to training is underway. The results of the initial research will serve as the guide for further R&D in the training area. If funding permits, plans call for continuation of the training applications through FY92.

Improving the Design of the Marine Corps Job Performance Measurement Project

Milton H. Maier and Paul W. Mayberry
Center for Naval Analyses

Overview

In 1981, the Marine Corps conducted a study to evaluate the feasibility of using hands-on tests of job performance to validate enlistment standards that are based on scores on the Armed Services Vocational Aptitude Battery (ASVAB). The results were favorable, and continued analysis of the data has provided new insights into the process of measuring job performance. This paper presents results from the feasibility study that were instrumental in changing the design of the Marine Corps project to measure job performance and validate enlistment standards.

Feasibility Study

Three occupational specialties were included in the feasibility study: Ground Radio Repair, Automotive Mechanic, and Infantry Rifleman. The Ground Radio Repair specialty has high technical demands (37 weeks of formal school training), the Automotive Mechanic specialty has moderate demands (13 weeks of training), and the Infantry Rifleman specialty has low demands (5 weeks of training). Marine Corps job experts, assisted by testing psychologists, developed a hands-on test and a written test for each specialty. Both tests were administered by noncommissioned officers to first-term Marines, plus a few Marines in their second term of enlistment. Only the hands-on tests are considered in this paper.

The validity of the ASVAB scores was evaluated by computing their correlation with the hands-on test scores. The validity coefficients were as follows:

Specialty	Validity coefficient[a]
Ground Radio Repair	.59
Automotive Mechanic	.56
Infantry Rifleman	.58

a. Coefficients are estimated population values of the appropriate ASVAB aptitude composite.

These validity coefficients are comparable to those obtained when ASVAB scores are evaluated as predictors of final grades in occupational specialty training courses.

A second finding was that training grades and hands-on test scores were reasonably well correlated. The estimated population values were as follows:

Specialty	Correlation coefficient
Ground Radio Repair	.52
Automotive Mechanic	.51
Infantry Rifleman	.39

The third finding was that ASVAB qualification standards derived from the feasibility study agree closely with the existing Marine Corps standards. The two sets of standards, which are expressed on a standard score scale with a population mean of 100 and standard deviation of 20, are as follows:

	Qualification standards	
Specialty	Existing	Derived
Ground Radio Repair	115	115
Automotive Mechanic	90	95
Infantry Rifleman	80	85

Based in part on these results, the Marine Corps continued its support of the Job Performance Measurement Project.

Effects of Job Experience on Job Performance

A second round of analyses examined the effects of job experience, measured as time in service (TIS), on the hands-on test scores. The intent was to evaluate the content validity of the hands-on tests. Performance was expected to increase with TIS. Another expectation was that people with high aptitude should improve more than people with low aptitude. The reasoning was based on a widespread definition of aptitude as the ability to profit from experience. In traditional training courses, with length of course held constant, people with high aptitude profit more, as measured by grades, than people with low aptitude. Extending the argument, people with high aptitude should become progressively better performers than people with lower aptitude.

The analysis showed, however, that the hands-on test scores increased only modestly, if at all, during the first term of enlistment. Because the infantry riflemen tended to have limited job experience, no clear trend was established for them.

The next step was to see how aptitude was related to the hands-on test scores for people with different levels of experience. The results for radio repairers and automotive mechanics are as follows:

Ground Radio Repair

Months in service	Mean test score	Validity coefficient[a]	Number of cases
15-25	48.5	.69	38
26-35	50.0	.00	53
36-48	52.6	.00	37

Automotive Mechanic

2-14	47.6	.72	57
15-25	50.0	.52	56
26-34	50.3	.15	53
35-60	53.0	-.07	54

a. Population-wide estimated validity of appropriate ASVAB aptitude composite.

The startling finding was the precipitous drop in the validity of the ASVAB for people with 2 years or more in the Marine Corps. After 2 years of experience, on average, people with low aptitude caught up to those with high aptitude, which resulted in little change in the hands-on test scores for the total sample and essentially no ASVAB validity for the more experienced people.

This finding raised questions about the appropriateness of the hands-on test for people with more than 2 years of experience. If the hands-on tests in the feasibility study were appropriate measures of job performance for all levels of job experience, then job experience benefits people with low aptitude, but not people with high aptitude. If the test content is appropriate and the ASVAB is found to have low validity for predicting job performance of experienced workers in other occupational specialties, then justifying ASVAB enlistment standards becomes more difficult.

An alternative view is that the test content was not appropriate for many people with more than 2 years of service. Some people with high aptitude may have had different job requirements from those covered by the hands-on tests, and then they could not demonstrate their true proficiency on the tests. If so, their mean hands-on test score would not increase with TIS, and the apparent validity of the ASVAB would drop.

These results, even though not definitive, had an impact on the design of the Marine Corps project.

The procedures for developing the tests were tightened to ensure that their content reflected job requirements at all levels of experience.

The composition of the sample of examinees was expanded. Instead of testing only first-termers at grades E-1 through E-4, the samples now include second-term Marines at grades E-5 and E-6. Two levels of the hands-on tests are being developed, one for first-termers and one for second-termers.

The sample size was expanded. The original design called for testing 300 people, with experience concentrated at about 2 years TIS. Now the sample size is 1,800 people, with 1,200 in their first enlistment and 600 in their second. With a sample of this size, and tests with appropriate content, the question of how well the ASVAB predicts job performance for experienced Marines can be answered quite definitely.

A bonus from this design is that the Marine Corps can evaluate reelistment standards as well as initial enlistment standards.

Effects of Test Administrators on the Validity of Hands-on Tests

The importance of test administrators for obtaining valid measures of job performance became more apparent as the analysis proceeded. Test administrators score the tests while administering them, and any differences in scoring standards reduce the measurement validity of the tests. If the administrators employ the same scoring standards, then means, variances, and intercorrelations for the tasks in the hands-on tests should be essentially equal.

For the radio repair and automotive mechanic sample, the test administrators in the feasibility study could be identified and their scoring standards compared. The administrators scored all tasks for each examinee they tested, which permitted an evaluation of task means, variances, and intercorrelations. For the infantry rifleman sample, the test administrators could not be identified and therefore they are not included in this analysis.

The means, variances, and intercorrelations differed significantly for both sets of administrators. Results are shown below for the task of maintaining wheels and brakes on the quarter-ton Jeep. Sixty minutes were allotted for this task, which had 13 steps that were scored as pass/fail.

Test administrator	Number of cases	Mean score	Standard deviation	Correlation with the coil test
1A[a]	55	52.76	14.06	-.06
1B	40	47.02	5.21	-.03
2	41	46.82	7.07	-.02
3	13	47.99	6.96	.56
4	27	54.47	7.06	.48
Total	176	50.00	10.00	.08

a. Administrator 1 tested 95 people, 55 in the first half of the testing period (1A), and 40 in the second half (1B).

The conclusion is that the test administrators did not use the same scoring standards. As a result the design of the Marine Corps project was changed as follows:

Test administrators are to be trained to use the same scoring standards.

Test administrators are to be monitored daily by profiling the scores assigned by each administrator to each examinee on each task. The mean profile for administrators, summed across examinees, should be essentially parallel. Deviations would suggest idiosyncratic standards.

Administrators are to be rotated among testing stations. Each administrator will be responsible for a set of tasks administered at

a testing station. To minimize systematic effects on examinees' scores, administrators will be rotated daily at random among the testing stations. This rotation scheme will enable an analysis of interaction effects between test administrators and testing stations. If the administrators are properly calibrated to the scoring standards, the interaction effects will be negligible. If there is an interaction effect, the quality control failed and data editing becomes more complex.

The effect of this redesign will be to produce data of sufficient quality and quantity to address questions about the validity of using ASVAB scores to select and classify Marine recruits. The current system has been in effect since World War II, and has been functioning reasonably effectively. But the paucity of data about the relationship between aptitude and job performance has not allowed satisfactory answers to the questions about the validity of the ASVAB. At the conclusion of this study, the status of aptitude tests in personnel decisions should be clear.

Display/Cognitive Interface:
The Role of Display Proximity in Information Integration
Tasks Relevant to Command, Control, and Communications (C3) Displays

Christopher D. Wickens, Melody Carswell, Lee Fracker,
Barbara Barnett, and Arthur Kramer
Institute of Aviation, Willard Airport
University of Illinois at Urbana-Champaign

Abstract

A series of three experiments are described that test the hypothesis that information sources which must be integrated should be displayed in close spatial or configural proximity to each other. The three experiments mimic different aspects of the information processing demands of C3 tasks. In Experiment 1 (dynamic process monitoring) and Experiment 3 (multi-cue decision-making), integrated polygon displays were compared with separated bargraph displays and found to be superior. In Experiment 2, examining target search, the display of information in common perceptual formats (codes and modalities) was found to be beneficial. Experiments 1 and 3 also examined the <u>cost</u> of integrated displays to the processing or memory of isolated stimulus dimensions.

Introduction

The computerized Command, Control, and Communications (C3) system has the potential to impose extremely heavy information processing loads on system monitors and supervisors--loads that can potentially exceed the human's ability to process all of the relevant information. Given the availability of flexible software and display interfaces, along with color and voice technology, it is critical for research to identify those characteristics of perception and cognition that will guide the choice of display technology to allow the most rapid integration of large volumes of information. There are, indeed, numerous principles of display design that may be brought to bear in effective display formatting: Many of these relate to sensory properties, addressing such characteristics of luminance, glare, visual angle, noise, and so forth (Van Cott & Kinkade, 1972; Boff & Kaufman, 1986). Less well-developed in this context are cognitive models that address the interface between the perception of information (as dictated by the display format), and the mental integration or understanding of that information; that is, the display/cognitive interface.

Three such principles have received some validation: (1) Where possible information should be displayed using <u>familiar</u> words, acronyms, icons or graphics. (2) When muliple channels must be processed <u>independently</u>, there is an advantage gained to using <u>separate</u> <u>resources</u> to display the task-relevant information (i.e., separate perceptual modalities, Wickens, 1984). The third important principle, which forms the basis for the three experiments reported here, is an extention of the general principle of <u>display</u>

compatibility. This extention, which we describe as compatibility of proximity proposes that, to the extent that 2 or more sources of information must be mentally integrated, there is an advantage to displaying that information in close proximity. Close proximity in this context may refer either to similar perceptual modalities or codes, to close proximity in space, or to the configuration of visual dimensions as properties of a single object. This principle has received some support in our previous research (e.g., Wickens, Boles, Tsang, & Carswell, 1984). Here we report three experiments that confirm its validity in different contexts, but also assess whether there is a cost to this display proximity, in terms of the operator's ability to process the stimulus channels independently. These experiments do not attempt to recreate the full complexity of C3 displays. However, an effort is made to simulate the kinds of information processing demands that characterized the C3 enviroment: Understanding cause and effect relations, locating computerized information in spatial maps, and integrating probablistic information in multicue decision making. Because of space limitations, the details of these experiments will not be reported here, and the reader is referred to Wickens et al., 1985 for further elaboration.

Experiment 1. Object Integrality in Dynamic System Monitoring

The monitor/supervisor of any complex data base may often be required to monitor the causal relations between system varibles, to establish if the system is functioning in an expected fashion. During the 1984 proceedings of the symposium, we reported an experiment in which subejcts monitored two simulated dynamic processes to detect failures in the mathematical relations between each system's two inputs and its output (Wickens et al, 1984). In that paper we reported that failure detection was significantly faster and more accurate when the input-output relationships of the two processes were displayed as the width of the base of a triangle and its height, than as three separate bargraphs (see Figure 1). That is, the integral object display better served this information integration task.

One condition of the present experiment replicated the information integration requirements of the previous study. However, in addition two further manipulations were inposed. In one of these -- the multiple task manipulation -- subjects were required to monitor each of the six variables independently, and press a corresponding button whenever each crossed a zero reference marker, as shown in Figure 1. The other manipulation was imposed on both the integration and multiple task condition. This was a secondary memory task. At unpredictable intervals during both tasks, the display would suddenly blank, and this was followed by a probe request for the subject to recall the value of one specific dimension. Therefore both the multiple task and the memory task demands differed distinctly for the original integration task, by requiring subjects to pay attention to the attributes in isolotion, rather than their integration. It was hypothesized that this request to focus or divide attention would shift the balance in favor of the more separate display. Ten undergraduates were assigned to either the integration or multiple task condition. In the integration condition two independently time-varying inputs were equally weighted and summed to produce the output. When the system failed, the contributions of these inputs to the output diminished to zero over a five second interval, and was replaced by a third randomly varying signal, of the same spectral characteristics. A failure thus was reflected in a shift in the correlational pattern between variables. Failures accured at random, unpredictable intervals during the 3 minute monitoring

trials. For the subjects in both groups, half of the trials imposed the memory probe task concurrently with either integration (for one group), or reference-crossing monitoring (for the other). One quarter of the trials involved only the memory task, and one quarter involved the primary task.

Full analysis of these data are reported in Wickens et al. (1986). In this discussion we shall only describe qualitatively the important principle effects. These are: (1) The beneficial effect of the triangle display over the bargraph display for the integration task was replicated. (2) This object-display advantage was not observed in the multiple task condition. In fact, under these circumstances, the separated bargraph display showed a non-significant performance advantage. (3) Concurrent performance of both the memory task and the primary task produced interference in both. However, the effect of display type on memory performance was different for the two primary task groups. When the integration task was performed, memory for individual attributes was better when those attributes were also displayed as separate bargraphs. Hence, while the object display was helpful for integration, it was disruptive for the memory of isolated attributes, just as it had been shown to be disruptive when the task required independent, rather than integrated processing of the separate attributes. The results of the memory task for the multiple task group are somewhat more complex and are described in Wickens et al. (1986).

Experiment 2. Target Location

In the 1984 paper we described an experiment in which integration of information was better served by homogeneous displays, made up of all verbal or of all spatial information, than by heterogeneous displays, combining the two codes of display format. In the present experiment we examine this effect of display proximity of code, and also of modality in a target location task. In the task, the subject is cued in advance as to the defining attributes (shape, color, location) of a target for which he must search. Then a visual analog display appears with the cued target appearing in one corner, and the subject must signal as rapidly as possible the location of that corner. The prior target information appears in one of three formats that systematically vary in their proximity to the subsequent visual-analog search display.

The prior information was either 3 spoken words (e.g., "left blue square" -- different modality and code), - 3 printed words in the middle of the square search array, with the same timing as the auditary display (same modality, different code), or three analog representations of the three attributes (same code and modality -- closest proximity). This information was also presented sequentially in the middle of the search array. Immediately after the termination of the last attribute of target information, the search display appeared with the target appearing in one of the 4 corners, and "foils" appearing in the other three. The subjects pressed one of four buttons to designate the corner in which the target appeared. Six subjects participated in the experiment which lasted for 3 one hour sessions.

The results of Experiment 2 were quite straight forward. The three conditions differed from each other only in terms of latency, which increased monotonically as the target information display differed from the search display by progressively more features: Visual Spatial - 570 msec, Visual Verbal - 690 msec, Auditory Verbal - 745 msec. Analysis of variance revealed a significant main effect of conditions ($P < .05$), while planned contrasts

showed that the visual spatial display differed from the two verbal conditions, but the latter two did not differ from each other. Hence the results replicated the findings of Wickens et al. (1984), regarding the role of processing code similarity in the proximity compatibility hypothesis. The role of modality (visual vs. auditory) similarity is suggestive, but in light of the marginal reliability of the effect it does not appear that any great advantage is gained by using common processing modalities.

Experiment 3.

A major component of C3 operations requires integrating probablistic information from multiple sources in order to formulate appropriate decisions. This probablistic characteristic means that the accuracy of the information in predicting the future or diagnosing the present state of the world is less than perfect. This accuracy may be dimished by two factors -- lack of diagnosticity -- an information source may simply not discriminate one state of the world from another -- and lack of reliability -- the source may not be trustworthy. Together these two quantities may be expressed as correlation coefficients and combined multiplicatively to produce the total information worth of a source -- a value that may of course range from 1 (=[r=1]x[D=1]) to 0 (if either R or D = 0).

In the present experiment we examine the ability of subjects to combine R & D information from each of several cues, bearing on one of the two predicted states of the world -- that a hypothesized air mission either will or will not be carried out successfully. Proximity of the R & D information and of the separate cues in this integration task, is varied in three respects. First, and most important for the present discussion, proximity of the R & D measures was varied by either presenting these as two adjacent bargraphs, or as the height and width of a rectangle. The latter technique of object integration developed by Scott & Wickens (1983) has the added benefit that the area of the rectangle is directly proportionate to information worth. In addition to this manipulation of proximity within a cue, proximity between cues was also varied in time (by presenting them simultaneously or sequentially), and in space (by presenting the sequential cues in one location or four). Finally speed stress was manipulated in the simultaneous cue presentation condition, by presenting the four cues for either 1 second or 4 seconds viewing.

The subjects were divided into two groups of four, each subject assigned to either the bargraph or rectangle display condition. Subjects then viewed a series of trials, each trial consisting of the sequential or simultaneous presentation of the four cues. On any given trial, these four were selected randomly from eight possible cues. Each cue was either a rectangle or a pair of bargraphs (depending upon the subjects' condition), and was colored white, if it supported the decision to carry out the mission, or black if it supported the decision to abort. The subject was to aggregate the collective evidence across the four cues, and express this as a confidence rating on a continuous bipolar scale. Performance was measured by the correlation between these ratings and the optimal confidence response over trials. Subjects performed the experiment over 4 one hour sessions. As in Experiment 1, on periodic trials subjects were probed as to the value of isolated dimensions.

The results indicated a clear superiority for the rectangle object display over the bargraph condition as measured by the actual-optimal correlations. However, unlike the results of Experiment 1, this condition

showed no evidence of a deficit in the subjects' ability to recall the isolated dimensions that were probed. Finally, subjects who performed with the rectangle display showed immunity to the effects of increased speed stress, while performance of those in the bargraph condition suffered accordingly. The effects of proximity of space and time between cues were less pronounced.

Discussion

The results of all three experiments, along with those reported in this volume by Kramer, et al., suggest that the compatability of proximity principle is a meaningful one that could and should have implications for the manner in which information displays are formatted. Of course, the paradigms reported here were meant to simulate the processing demands of the C3 task, but neither the informational complexity, nor the operational stress of the C3 environment. While data from such an environment remains still to be collected, there is little reason to doubt that the principles validated here will continue to apply (Wickens, et al., 1985).

Acknowledgments. The research reported in this paper was supported by Contract No. F30602-81-C-0206 (B-5-3705) from the Rome Air Development Center. The authors would like to acknowledge the contribution of Roger Marsh, Gunnar Seaburg, and Todd Muslech for programming assistance, and Lisa Weinstein and Keith Glasch for assistance in subject running.

References

Boff & Kaufman (1986) Handbook of Perception & Performance: Wiley.

Scott & Wickens (1983) In 27th Proceedings Human Factors Society Meeting.

Van Cott & Kinkade (1972) Human Engineering Guide to Equipment Design. U.S. Govt. P.O.

Wickens (1984) Engineering Psychology & Human Performance: Charles Merrill.

Wickens, Boles, Tsang & Carswell, (1984) In 9th Psychology in DOD Proceedings.

Wickens et al (1986) Display/ Cognitive Interface. U. of Ill., Eng. Psych. Lab. Tech. Rep. EPL 85-2/RADC 85-1

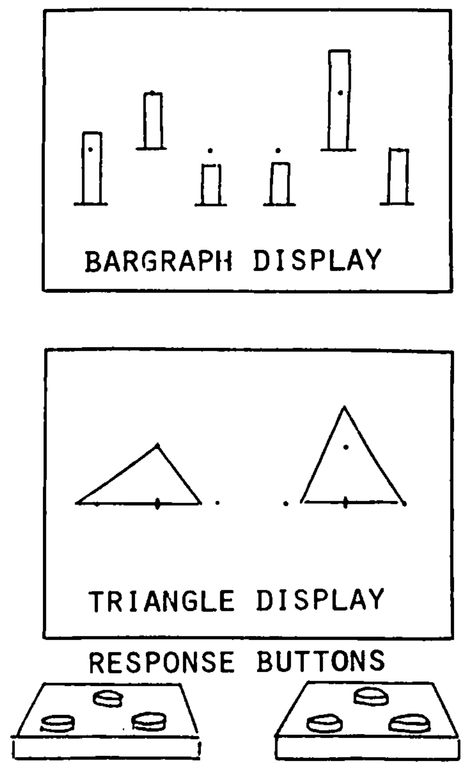


Figure 1. Displays used for experiment 1.

Extraction and Comprehension of Data Relations: An Application of the Display Proximity Principle

Arthur Kramer, Christopher Wickens, Barry Goettl & Kelly Harwood
Institute of Aviation, University of Illinois

Abstract

Three experiments were performed to determine whether tasks that require the integration of several channels of information would benefit from displays which represented this information in close proximity. We define "display proximity" in terms of an advantage in multi-task processing obtained by representing task parameters by different levels of a dimension or different attributes of an object. Another objective of the studies was to determine whether there was a cost to display proximity when the tasks required selective processing rather than integration. In general, the results of the experiments support our hypotheses. Tasks which require the integration of several channels of information benefit from close display proximity while tasks that require selective processing are better served by separate display representations.

In recent years, the human operator has been increasingly called upon to serve as a data manager/supervisor of complex, multidimensional systems. A case in point is the computerized Command, Control and Communications (C3) system in which enormous amounts of data must be encoded, processed and acted upon by the operator in relatively brief periods of time. Thus, one of the important design considerations for these systems is the configuration of displays in a manner which is consistent with the capabilities and limitations of human information processing. However, few guidelines exist for the design of the interface between the multivariate database which forms the basis for systems such as C3, and the operators' comprehension and integration of the relevant information into a dynamic mental model of the system.

Multiple Resource theory has been quite successful in providing display formating guidelines for a limited subset of tasks (Wickens, 1980). This theory proposes that the human operator possesses several distinct types of processing resources which may be shared among concurrently performed tasks. These resources include, stages of processing (perceptual/central vs response), codes of processing (spatial vs verbal), and modalities of processing (visual vs auditory). The efficiency of multiple task performance has been shown to be a function of the degree to which tasks require the same resources, with higher demands for common resources leading to poorer performance. In terms of display guidelines, these findings suggest that displays should be configured so as to represent concurrently performed tasks as "far apart" as possible in the resource space. For example, efficiency will improve if one task is displayed visually in a graphic format and the other task is represented auditorially (Wickens, Sandry & Vidulich, 1983).

However, subsequent research has shown that the benefit derived from using separate resources for multi-task performance does not hold when the information must be integrated into a single mental model of the task (Carswell & Wickens, 1984). In this case, performance improves when different channels of information can be processed within the same resource space. Thus, for tasks such as C3 which require that the human operator make decisions on the basis of the overall picture of the system, displays should be designed to represent the relevant information in close "proximity". The relationship between the distance of tasks in the resource space and the processing requirements of separate and integral tasks has been referred to as the "principle of display proximity" (Wickens & Boles, 1983).

The concept of display proximity can also be defined at other levels of analyses. For example, the distance among concurrently performed tasks might be described in terms of spatial and semantic dimensions. Proximity may also be defined with respect to the representation of system parameters as multiple attributes of a single object (i.e. color, shape, size) or as multiple objects. Recent research has indicated that attributes of a single object tend to be processed *in parallel* while separate objects appear to compete for a limited quantity of resources (Duncan, 1984; Kahneman & Henik, 1981; Kramer, Wickens & Donchin, 1985).

In the present paper we report the preliminary results of three studies which examined the concept of display proximity in terms of object integration. In all cases, subjects were required to integrate several channels of information into a single mental model of a system. The first

two experiments focus on data comprehension in static systems while the third study investigates the effects of display integration on monitoring and failure detection in a dynamic information managment system.

Experiment 1 - Data Extrapolation

Method The purpose of this study was to test the hypothesis that perception of the relations between data will be facilitated by graphical displays that integrate data points together via object configurations. Twenty two subjects were presented with speed and accuracy measures for two conditions in a hypothetical experiment. The subjects were then presented with the quantitative value of the independent variable for a third condition, and were asked to estimate its speed and accuracy from their understanding of the functional relationship. Two different displays were used to present the data. In one display the speed and accuracy values for each of the two hypothetical conditions were represented by a single point on a coordinate graph. Thus, for this display condition subjects were instructed to indicate the speed and accuracy values of the third condition by drawing a point on a blank coordinate axis graph. In order to equate response complexity between the display conditions, subjects were asked to extend lines from the point to the two axes. In the other display, the speed and accuracy values for each of the hypothetical conditions were represented by separate bars. Subjects were instructed to draw lines on the graph to represent the speed and accuracy of the extrapolated condition. Subjects performed the task with displays which indicated both positive and negative relations (slopes) between the two hypothetical conditions. Since the task was self paced, accuracy of extrapolation served as the performance measure.

The inherent relation among the data is such that each hypothetical condition is represented by two variables. Therefore, it was predicted that the coordinate graphs would result in more accurate extrapolation since they better conform to this inherent relationship than the bargraphs. Furthermore, it was predicted that performance would benefit in the coordinate graph by virtue of the fact that the relation between variables is represented by two objects, rather than four, in the bargraph display.

Results and Discussion The results confirm the hypothesis that coordinate axis graphs result in more accurate performance than the bargraphs (all reported effects are significant at $p < .05$). Two compatible reasons for this superiority may be offered. The first pertains to simple object superiority in this integration task. Since latency and accuracy of two points must be integrated to infer their relation and thereby extrapolate to the third point, the original hypothesis predicts that the integration will be better supported by a single object (point) than by two bargraphs. The second possible reason, relates to the inherent nature of the data. The assumed relationship is that of independent variables to dependent variables. It is presumed that subjects conceptualize the data in terms of an independent variable (two conditions) each having two dependent measures (speed and accuracy). The bargraph display is incompatible with this conceptualization of the data since it tends to emphasize the dependent measures, rather than the independent variable.

The coordinate graph is more compatible with the assumed relationship among the data, and apparently lends itself to simultaneous extrapolation. In contrast, the bargraph seemed to force a serial process of extrapolating speed and then accuracy. This sequential characteristic was supported by the large effect of variable in bargraph displays relative to that observed in coordinate graphs.

Experiment 2 - Multi-Cue Judgment

Method This experiment examined the effects of display format on learning and performance in a multi-cue judgment (MCJ) task. Subjects were presented with three cue values and were asked to predict a criterion value based on the cues. The subjects' goal was to use the feedback concerning the accuracy of their prediction to discover the underlying algorithm and improve performance.

Twenty four subjects participated in two experimental sessions. Half of the subjects performed the task using a bargraph display while the other half used a triangle display. The three cue variables were designated along the horizontal axis of the displays, while the vertical axis indicated the values of the three variables (ranging from 1 to 10). The heights of three bars represented their cue values. In the triangle display, the three points were connected to

form a triangle. An additive formula (criterion = (X+Y+Z)*4) was used to relate the cue values to the criterion.

Three blocks of 120 trials were run in each of the two sessions, with the first block in each session treated as practice trials. In the first session, the subjects saw a display of the cue values and then entered their estimate of the criterion. The computer then displayed the correct criterion and the subjects' response. The same procedure was followed for session two except that one of the three variables in each equation became irrelevant for an entire block of trials. The irrelevant variable was selected at random for each block and was not used in the calculation of the criterion.

The MCJ task requires the integration of information, thus to the extent that the display aids the subject in this process, performance should be improved. Therefore, it was predicted that in Session 1 the triangle display would lead to better performance than the bargraph. On the other hand, when the task requires the subject to selectively attend to a subset of the displayed variables as was the case in Session 2, the holistic processing engendered by the integrated display may actually degrade performance. In this case, the separable bargraph display is relatively more compatible with the relations among variables and thus should lead to superior performance.

Table 1: Z Scores for the Multi-Cue Judgment Task

	Session 1		Session 2	
	Block 2	Block 3	Block 2	Block 3
Triangles	1.30	1.45	1.60	1.85
Bars	1.35	1.60	2.80	3.00

Results and Discussion The correlations between the actual criterion values and the subjects' predictions were transformed into z-scores using the Fisher procedure: these were then subjected to three-way (display x session x block) mixed factors ANOVA. Overall, the results of this experiment provided partial support for our predictions of an interaction between display format and data relations. Previous studies had found a superiority of integrated displays in the learning of multi-cue relations (Goldsmith & Schvaneveldt, 1984). In the first session, performance in the two display conditions was not significantly different (see Table 1). However, in the second session when the task required subjects to ignore one of the three cue variables, performance was limited in the integral display, whereas performance continued to improve for the bargraph display. It should be noted that the separable superiority effect was demonstrated by an increment for the separable bargraph display rather than a decrement for the integrated triangle display. There are two potential explanations for this; (1) practice effects and (2) computational complexity. The latter can be elaborated by pointing out that when one variable becomes irrelevant, the subjects were required to add two numbers rather than three. In this way, any increased difficulty in attending to two cues in the integral display may have been offset by an overall decrease in computational difficulty.

Experiment 3 - Exploring and Understanding Multi-dimensional Data Bases

Method This study focuses on subjects' ability to identify the relation between the attributes of several "objects" or units. The experimental task required subjects to monitor multiple dynamic units with common variables (e.g. multiple attributes). In order to assess how well the relationship between and within units was understood, periodic "failures" were imposed into the data base. The subject's task was to identify the failures among different items within as well as between units. The display of the information was varied orthogonally to emphasize proximity within a unit, or proximity of similar attributes across units. The use of color along with spatial position was examined as a means of increasing proximity on both dimensions simultaneously (within and across units).

A bargraph format was used to display information. Each bar corresponded to one of the four attributes. The heights of the bars were updated discretely every 10-15 seconds across trials, in increments ranging from one to ten arbitrary units. Subjects decided if a failure had occurred based on their memory of the height of the attributes on the preceding trial. Accuracy and

reaction time served as dependent variables. Two different spatial display formats were used:

1) Two separate objects - attributes of each unit were grouped in two separate, spatial locations; specifically one graph above the other.

2) Single object - all eight attributes of both units were grouped together in one spatial location. Specifically, the same attribute from both units was paired together, thereby emphasizing proximity within an attribute.

The first format emphasized the "unit" quality of displayed information by clustering all attributes of one unit together, while the second format emphasized the "attribute" characteristics of the information. To further vary the nature of the relationship between attributes of the objects, three color coding schemes were chosen:

1) All attributes for both units were represented in the same color. Thus, the relationship between attributes was identified purely in spatial terms,

2) The attributes of each unit were coded by a unique color. This emphasized attribute relationships within an object.

3) The same attributes across both units were the same color. This coding emphasized attribute relationships between objects.

Thus by manipulating spatial position and color coding, a means was afforded for varying proximity on both the attribute and object dimensions simultaneously.

Since the present task requires the integration of information both across and within objects, we predict clear advantages for color and spatial coding. We also expect that poor grouping on one dimension (color or space) could be compensated for by emphasizing grouping on the other dimension. Furthermore, we predict that events involving a change in relations between attributes within a unit will be more easily detected than events that change attributes between two units because the object representation of a unit will increase the parallel processing of its attributes.

Table 2: Response Time in secs and (error rate)

	Object Emphasis (Dual Locations)			Attribute Emphasis (Single Location)		
	Normal	Between	Within	Normal	Between	Within
Mono-	11.57	10.04	11.00	12.24	11.14	12.66
chrome	(.27)	(.18)	(.27)	(.27)	(.41)	(.38)
Colored	11.36	9.62	9.88	10.60	9.61	9.91
by objects	(.38)	(.31)	(.31)	(.33)	(.35)	(.30)
Colored	9.85	9.22	10.51	10.13	10.00	12.07
by attributes	(.27)	(.31)	(.29)	(.29)	(.35)	(.33)

Results and Discussion The results confirmed our prediction that physical grouping of attributes with their respective objects would aid performance. This effect was supported by both accuracy and latency measures. Color coding also showed the same benefits as spatial proximity. There was a clear advantage for the color conditions over monochrome for the latency measure. However, the primary advantage here was of the two color conditions over the monochrome. The slight, 130 msec advantage to coloring by objects was in the right direction, but was not significant. We also confirmed our hypothesis that poor grouping by attribute or unit on one dimension (space or color) could be compensated for by emphasizing grouping on the other dimension. Thus, when objects were grouped by space, performance was improved when attributes were grouped by color. On the other hand, when objects were grouped by color, performance was helped when attributes were grouped by space. One surprising effect was the advantage of detecting between object failures over within object failures. We had reasoned that within object failures would be easier to detect because of the potential for parallel processing of the attributes of an object. A potential explanation for this effect is the possibility that since between object failures were more difficult to detect subjects emphasized them at the expense of the within object failures. The examination of this hypothesis will await replication of the present experiment using an experimental design in which failures are varied between blocks, rather than randomized within blocks of trials.

General Conclusions

The results of the three experiments generally support predictions derived from the principle of display proximity. This principle suggests that tasks which require the integration of several channels of information will benefit from displays which represent the relevant variables in close proximity while tasks that require selective processing will be better served by separate representations of task relevant information. In the present series of experiments, close proximity was defined in terms of displays which represented relevant information as different attributes of a single object rather than different objects and as different levels of a dimension (e.g. color, space) instead of different dimensions.

The results of Experiment 1 suggest that the understanding of multidimensional data is enhanced when this information is conveyed in an integral display format. This was accomplished by representing two dependent variables, speed and accuracy, as a single point in a corrdinate axis graph. Subjects were more accurate when extrapolating from two speed/accuracy conditions to a third condition when the data were displayed in a coordinate axis plot than when they were represented as a bargraph, with separate bars for speed and accuracy. Additional support for the facilitative effects of integral displays was provided by Experiment 3. In this case, the beneficial role of proximity of both space, through organizational grouping, and color, through the integration of information about units and attributes, was demonstrated. Furthermore, poor grouping by attribute or unit on one dimension (space or color) could be compensated for by emphasizing grouping on the other dimension. Experiment 2 also examined the potential cost to integral display formats when subjects were required to focus attention on certain elements while ignoring others. In this case, prediction of the criterion value in the MCJ task was more accurate when the cues were displayed as three separate bars instead of vertices of a triangle.

In conclusion, the results of this research indicate that the display proximity principle is a valid one, at least in a limited set of experimental paradigms. However, it is apparent that the true applicability of this principle to the C3 system awaits additional validation. The paradigms employed here were selected to mimic the information processing demands of different C3 tasks, and not the informational complexity or the stress of the operational environment.

Acknowledgments

This research was supported by a contract from Rome Air Development Command, F30602-81-C-0206. The authors would like to thank Gunnar Seaburg for programming assistance and Yili Liu and Lisa Weinstein for their help in running subjects.

References

Carswell, M. & Wickens, C. (1984). Stimulus integrality in displays of system input-output relationships. Proceedings of the 28th Annual Meeting of the Human Factors Society. CA.

Duncan, J. (1984). Selective attention and the organization of visual attention. Journal of Experimental Psychology: General, 113, 501-517.

Goldsmith, T. & Schvanveldt, R. Facilitating multi-cue judgments with integral information displays. In J. Thomas & M. Schneider (Eds.), Human Factors in Computer Systems. NJ: Ablex.

Kahneman, D. & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. Pomerantz (Eds.), Perceptual Organization. Hillsdale, NJ: Erlbaum.

Kramer, A., Wickens, C. & Donchin, E. (1985). Processing of stimulus properties: Evidence for dual task integrality. Journal of Experimental Psychology: Human Perception and Performance, 11, 393-408.

Wickens, C. (1980). The structure of attentional resources. In R. Nickerson (Ed.), Attention and Performance VIII. Hillsdale, NJ: Erlbaum.

Wickens, C. & Boles, D. (1983). The limits of multiple resource theory: The role of task correlation/integration in optimal display formatting. Engineering Psychology Research Laboratory, University of Illinois, Technical Report EPL-83-5/ONR-83-5.

Wickens, C., Sandry, D. & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing and output. Human Factors, 25, 227-248.

Differential effects of memory load on the performance of substantively different aspects of a computer game.

David B. Porter

Oxford University

A rapidly-paced, visuo-manual computer game composed of substantively different subtasks (one uncertain and the other complex) was combined with an auditory-verbal side task involving memory loads of 0, 3, or 5 letters. After independently practicing the game, ten male and ten female paid subjects performed each of nine possible memory load and subtask priority combinations twice. Increased memory load significantly reduced performance on the uncertain task. In contrast, performance on the complex task actually improved with increased memory load. Analysis of post-task questionnaires revealed that although espoused strategies were positively related to performance of the uncertain task, espoused strategies for the complex task were inconsistent with observed performance. These results suggest the utility of explicit knowledge (and the cognitive resources which employ such knowledge) critically depends on the type of task and validity of the knowledge.

An experiment by Brown (et.al., 1969) is a noteworthy exception to the tradition of investigating human decision-making and performance separately. In this experiment, drivers negotiated a track containing a number of obstacles forming "gaps" while simultaneously answering questions. About half the gaps were slightly more narrow than the automobile; drivers were to detour around these "impossible" gaps. Thus drivers' performance in maneuvering the car through possible gaps and their decisions in choosing which gaps were possible provided independent measures of performance. The verbal side task caused considerable interference with decision-making but left driving performance unimpaired.

Brown's experiment introduces an important issue: the role of verbalization in integrated task performance. Although it has been widely assumed that the provision of more cognitive "resources" will enhance performance, recent findings suggest that for certain tasks inappropriate employment of explicit verbal strategies (by processing resources) can limit both current problem solving and subsequent recall (e.g., Reber, 1976 or Berry and Broadbent, 1984). A computer game which combined two substantively different tasks (i.e., a complex, certain, non-salient one and a simple, uncertain, salient one) was employed to seek evidence of differential effects of a resource-consuming verbal side task. To the extent subvocal verbalization enhances task performance, a verbal side task should interfere. If, however, verbalization is irrelevant or impedimentary, then the verbal side task should either have no effect or even facilitate performance. To the extent that a task is non-salient and, as a result, inaccurate or inappropriate explicit verbal strategies are developed, the latter rather than the former pattern should obtain.

PROCEDURES

Ten male and ten female paid subjects between the ages of 18 and 38 from the Oxford Subject Panel individually participated in this three-hour experiment. A 48 K Sinclair Spectrum microcomputer with printer and color monitor was employed for pregame testing and training as well as game presentation.

Subjects completed 150 trials of a four-choice reaction task. The index and middle fingers of both hands were overlapped on the keyboard to produce a direct spatial mapping for responses. Subjects next received training in controlling the movements of a blue whale (the protagonist of the game). Although the controls were the same as in the reaction task, subjects were required to hold the key down until the whale responded to compensate for a variable lag in the system. Subjects received as many practice trials as necessary to reach the control criterion (10 laps around an iceberg in less than 70 seconds). Some subjects reached the criterion on their initial trial while others required up to 7 trials. Total time in pregame

practice thus varied from 10 to 25 minutes depending on each subject's skill.

Subjects then completed 9 practice trials of the Save The Whale computer game. After a short break, subjects were taught to perform the verbal side task and then completed two additional sets of 9 trials combining different game prioritizations with levels of the side task. Since understanding the results of the experiment relies heavily on understanding the computer game itself, this will be described in some detail.

The monitor screen displayed a 21 X 31 cyan matrix containing four clusters of white icebergs as shown in Figure 1. Subjects controlled the direction of movement of a whale (initialized at screen center of each trial) by using the keyboard controls employed in practice. The whale had two substantively different tasks: eating plankton and wrecking kayaks. The differences between these tasks are important.

The plankton task was simple but uncertain. A single mass of plankton was initialized in one of four locations to the right of the whale. It proceeded in a diagonal psuedo-random walk to the right, drifting slowly to the top and then the bottom of the screen. It disappeared from the right screen border to reappear at the extreme left screen border. While its general path was repeated the specific sequence of its movement was generated by a complex formula and was unpredictable. The task was particularly frustrating because the plankton occupied only one space and the whale moved one space each cycle (ergo, it was impossible for the whale to move into the plankton's space half of the time). It was relatively obvious how to hit the plankton (i.e., by always turning towards it), but accomplishing this required considerable effort.

In contrast, the kayak task was more complex but completely predictable. Kayaks were "generated" at one of five locations along the screen border ("KG" in Fig. 1) at fixed times during the game. Once generated, all kayaks followed the same rule: they moved toward the whale. If they were either horizontally or vertically aligned with the whale, they moved one space horizontally or vertically (e.g., kayak A in Fig. 1). If not aligned, they moved one space vertically and horizontally (i.e., diagonally as would kayak B). Kayaks remained on the screen until one of two things happened: they encountered an iceberg and sank (as is about to happen to Kayak B) or reached the whale and harpooned it. Subjects gained points for the former and lost an equal number for the latter. The number of kayaks present at any one time varied from zero to four, but approximately 19 were presented during the course of each trial.

A trial lasted 2 minutes and 35 seconds and consisted of 218 "cycles" during which each of the characters could move one space. The two tasks were combined in three ways. In plankton priority trials, plankton was worth 100 points a bite and crashing or being harpooned by kayaks was worth plus or minus 10 points. In the kayak priority trials the points were reversed. In equal priority trials all contingencies were worth 50 points. The priority was explained by instructions preceeding each trial. Half the subjects started with a kayak priority trial first, followed by an equal priority trial and then a plankton priority trial. The other subjects accomplished the trial priorities in the reverse order.

Having completed 9 practice trials (3 of each priority in rotation), subjects received side task instruction. The side task involved subvocal rehearsal and report of random strings of either 0, 3 or 5 letters. The control condition involved a single consonant, and the two levels of memory load contained 3 or 5 non-repeated heterophonic consonants. The experimenter first announced the string, then 25 seconds later tapped a pen on the table for the subject to verbally respond. After the subject's response the experimenter replied "correct" (if appropriate), "close" (if a single letter was incorrect or transposed), or "incorrect" (if there was more than a single error). Subjects were requested to maintain correct performance throughout. In the control condition, the experimenter repeated a single consonant six times then immediately signalled for the response.

Each subject completed each of the nine possible combinations of memory load level and task priority twice in a Latin Square design. Order of presentation was counterbalanced across subjects. At the end of each trial subjects recorded their

scores on both subtasks from a monitor display and were told their combined score on the side task by the experimenter (1 for each "correct" string, 0 for each "close", and -1 for each "incorrect").

After the final trial, subjects were asked to rate several explicit rules which might apply to the two tasks. Subjects first indicated whether statements were something that should be done (+), avoided (-), or didn't matter (0). They then ranked the statements in order of importance. This was accomplished first for the plankton-eating task, then separately for the kayak-wrecking task. Subjects were debriefed and paid. Although most subjects required nearly 3 hours, they showed little sign of fatigue and many commented they had enjoyed playing the game.

RESULTS

The results will be presented in several parts. Means and standard deviations reflect the general level of performance. The most important data, however, are those which show the effects of the different conditions. Raw measures were "standardized" for each subject across all 27 trials. Standardizing scores permits a meaningful comparison of performance on the two tasks and equalizes the influence of each subject. Otherwise, subjects with greater variance have a disproportionately large effect on subsequent analyses. Standardization insures the each subject's performance receives equal weight. Multiple regression analyses with "adjustments to degrees of freedom" as suggested by Cohen and Cohen (1983) was employed to analyze the standardized within-subject variance. These data were also plotted (Fig. 2) to illustrate the effects. Finally, simple correlations were used to derive the "rules" reflected by regularities in the subjects' performance and then compared to subjects' mean explicit ratings of those rules.

Subjects were capable of performing each of the tasks. Overall, subjects changed the direction of the whales travel 74.8 (SD=16.4) times per trial (i.e., on 34% of the 218 cycles or about every 2 seconds). Their whales consumed an average of 19.4 (SD=19.0) tons of plankton and wrecked 9.9 (SD=3.8) kayaks per trial. When plankton was the priority the mean tons of plankton eaten was 36.0 (SD=14.6). Similarly, when the priority was on kayaks, the average number of crashes rose to 12.9 (SD=2.7). In other words, on average, subjects were capable of "scoring" on plankton about 17% of the time or wrecking about two thirds of the 19 kayaks launched per trial. Although performance on the side task was strongly influenced by the level of the memory load and decreased slightly with practice, there was no significant difference between subtask prioritizations.

The following multiple regression equations show the main effects (as partial regression coefficients) of the priority instructions (PRI), practice (RUN), and memory load (MLD) on the two criteria: plankton eaten (PE) and kayaks destroyed (KD). Criteria scores were standardized to means of 50 and standard deviations of 10. Variables reflecting potential interactions were examined, found to be non-significant and omitted from the equations as suggested by Cohen and Cohen (1983). Corresponding t values are shown below each of the predictors. The R^2, degrees of freedom, and confidence level for each equation is then listed.

PE =	12.07 PRI +	.57 RUN −	.54 MLD +	24.26
t(df=288)	23.79**	2.35*	−2.67**	
R^2 = .667	F(54,288) = 10.68	$p<.01$		

KD =	−8.60 PRI +	.60 RUN	+ .13 MLD +	61.74
t(df=288)	−11.15**	1.62	.43	
R^2 = .306	F(54,288) = 2.35	$p<.01$		

Priority (coded 1 for kayak, 2 for equal, and 3 for plankton) was by far the most important predictor of both criteria. Practice (RUN) (coded 4 through 9) was a positive predictor of performance on both criteria but was significant for only the plankton task (failure for a similar effect size to be equally significant for both tasks is due to the greater proportion of the total variance explained for the

plankton task). Memory load (coded 0, 3, or 5) was a significant negative predictor of performance on the plankton task. Although slight, the effect of memory load on the kayak task was facilitory.

Narrowing our attention to the performance of the two tasks when each was afforded top priority, allows simple graphical comparison across the three memory loads (Figure 2). The elevation of the plankton curve above the kayak curve suggests that of the total variance in the plankton task a larger portion was explained by priority instructions. This is consistent with both the larger partial regression coefficient and the suggestion that the plankton task was more obvious or "salient". The differential effect of memory load is apparent in Figure 2. However, direct analysis showed the apparent interaction between task type and memory load not to be statistically significant.

A number of other aspects of subjects' performance were collected during each trial. The simple (zero order) correlations between these subsidiary measures and each respective criteria provides an indication of their relative utility and are shown as hatched bars in Figure 3. These activities could also be verbally expressed as explicit "rules" to be followed in accomplishing each task. The plankton rules (and their within-subjects correlation with the criterion) were: RP1-Always turn toward the plankton (.91), RP2-Stay near the plankton (.57), RP3-Don't eat icebergs (.35), and RP4-Ignore kayaks (.23). The corresponding rules for the kayak crashing task were: RK1-Don't turn away from kayaks (.47), RK2-Eat icebergs (.43), RK3-Ignore the plankton (.23) and RK4-Stay near the center (-.07). Subjects' mean explicit ratings of whether these rules should be followed or avoided and their espoused relative importance is shown in Fig. 3. For the plankton task, the espoused rules correspond well to subjects actual performance of the task. However, for the kayak task the mismatch between the espoused and displayed rules is considerable. In fact, those rules most closely related to kayak-wrecking (RK1 and RK2) were espoused to be the activities most important to avoid.

DISCUSSION

The verbal memory side task showed different effects on the two subtasks as predicted. The significant decrement in the performance of the simple but uncertain (plankton) task caused by the side task, suggests a positive contribution of verbalization. The non-significant facilitation of the complex but certain (kayak) task by the same side task suggests that verbalization did not contribute to this task. The different validity of subject's explicit strategies for the two tasks provides a partial explanation. The rules subjects espoused were consistent with their accomplishment of the plankton task. However, the espoused rules for the kayak task were inconsistent with actual performance.

These results are interesting from a number of perspectives. They suggest that computer games offer an alternative for investigating the performance of tasks which require both problem solving and motor skills. There is also an interesting comparison between these results and those from Brown's (et.al., 1969) experiment. Tasks that showed interference (viz., deciding on "possible" gaps and tracking plankton) and those that were resistant to the verbal side tasks (viz., steering the automobile and wrecking kayaks) don't share many immediately apparent similarities (e.g., decisional level, relative difficulty, required frequency of environmental monitoring). There are, however, similarities within these pairs of tasks on the dimensions of complexity and uncertainty. Both tasks which showed interference were relative simple but uncertain (i.e., adequate performance of either required a single correct decision but subjects could predict neither the occurence of an impossible gap nor the plankton's next move). Although the activities involved in steering the car or wrecking the kayaks are both relatively complex, they can be represented by constant relationships and are thus predictable. These tasks showed no interference from the verbal side task.

It all depends...

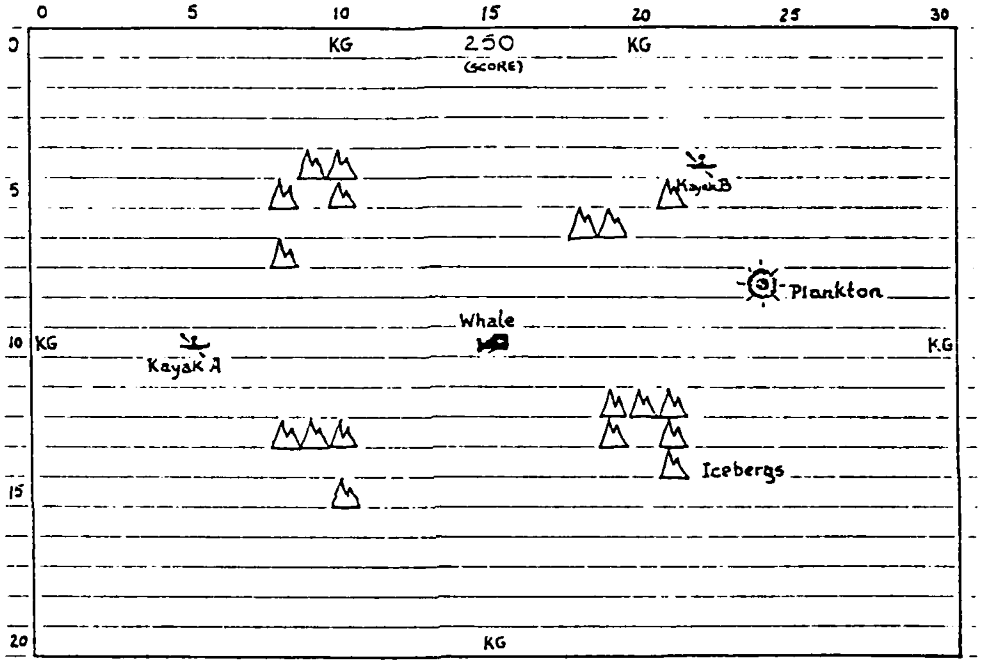
REFERENCES

Berry, D.C., & Broadbent, D.E. (1984). On the relationship between task performance and associated verbalizable knowledge. The Quarterly Journal of Experimental Psychology, 36A, 209-231.

Brown, I.D., Tickner, A.H. & Simmonds, D. (1969). Interference between concurrent tasks of driving and telephoning. Journal of Applied Psychology, 53, 419-424.

Cohen, J., & Cohen, P. (1983). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Hillsdale, NJ: Erlbaum.

Reber, A.S. (1976). Implicit learning of synthetic languages: The role of instructional set. Journal of Experimental Psychology: Human Learning and Memory. 2, 1, 88-94.



Monitor Display - SAVE THE WHALE Computer Game Fig 1

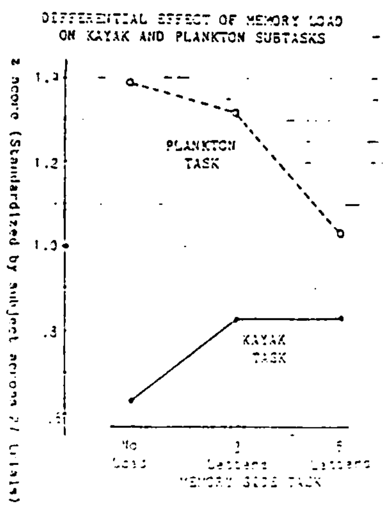
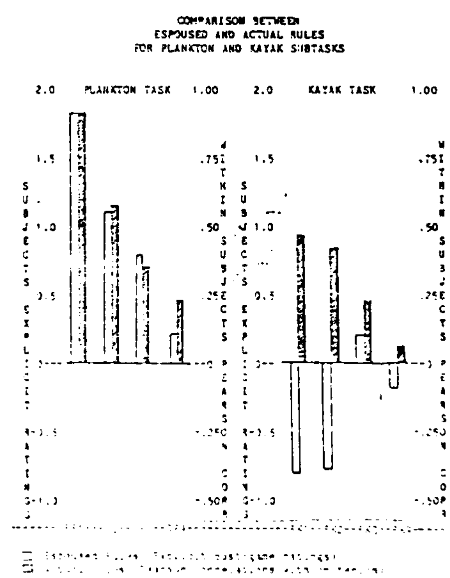


Fig 2

Fig 3 →



Analytical Studies of Recognition Memory

Delos D. Wickens
Colorado State University

Abstract

This research paradigm uses James's differentiation between Primary and Secondary memory, where PM refers to information still in consciousness and SM to material "of which meantime we have not been thinking" (working vs. inactive memory today). The procedure uses the Sternberg paradigm under these two conditions: PM wherein the probe follows the memory set immediately; SM where a distractor task is interpolated between memory set and probe. In SM the set must be retrieved to complete the decision task. Since otherwise the two tasks are the same; subtracting the PM score from SM gives a measure of retrieval time. Both conditions occur in the real world.

This paper reports several researches using the paradigm to measure retrieval time as a function of: Age (preschool, college, 70 year olds); repetition, interference, set size, and correlation with scan rate. Finally, the various results throw light on the nature of the retrieval act.

Introduction

Background

The paradigm to be described was developed as an analytical tool for studying memory processes. It uses the Sternberg (1969) procedure under the two memory conditions described by James (1896), namely Primary (PM) and Secondary (SM) memory. For James "an object of primary memory...[is not] lost from consciousness...[but] comes to us as belonging to the rearward portion of the present space of time...[However] secondary memory...is the knowledge of a former stage of mind..of which meantime we have not been thinking" (pp. 647-648). Most research with the Sternberg has been in PM, with the probe following the memory sets termination after an unfilled interval of about a second. In addition to using this PM procedure, Wickens, Moody and Dow (1981) created an SM procedure by introducing a distractor task of counting backwards by 3s for 12 sec between the memory set offset and probe onset. This required the subject to retrieve the memory set from SM in order to complete the scanning-judgmental task. Using the logic of Donders (1969), one can measure retrieval time by subtracting the PM score from the SM score. Both types of situations are found in the real world, where distractions from the task at hand are prone to occur, but the individual is expected to complete the task once the distraction is terminated.

Experiments

In this section the results of several experiments using this two-memory paradigm will be briefly described with the purpose of indicating the potential utility of the paradigm for the military.

Experiment 1: The Two-Memory Paradigm

The Wickens et al. (1981) experiment was originally devised to determine the locus of the proactive interference effect, whether at the encoding or retrieval stage. It used two groups of 48 subjects each, one in the PM condition, the other SM. The memory sets were 2 or 4 words, all words in a set coming from the same category. Three consecutive trials on one category were followed without interruption by a shift to another category and so on for 72 trials. The negative probe came from the current category. The interference effect was the difference between the first trial on a category and the third.

A regression analysis was conducted on those variables or interactions found to be significant in an analysis of variance, and the following equation offered the best fit to the data:

$$RT = [a + bm + t] + [(r_u + r_i) + q]$$

The first major term is all that was required to handle the PM data and is basically what Sternberg (1969) used. We substituted the symbol t for his symbol r to indicate the potential importance of the word used as a probe. Typically, and also in Wickens et al.'s experiment, the latency to the positive probe is less than to the negative probe. The term bm is, of course, the scan rate.

The second major term is a general retrieval term that was required to account for secondary memory performance. In it, r_u represents retrieval time, and r_i the increment produced by interference resulting from having experienced two prior memory sets from the same taxonomic category. The final term, q, was introduced to account for the fact that a marginally significant interaction (.07) between memory type and probe type was found. The rationale for its use is that the probe can serve not only as a signal to recall but also as a retrieval cue and that a member of the set itself is a better retrieval cue for that set than is a nonmember word.

An important finding in the data was the fact that the RT curves connecting the two- and four-item memory set were parallel--differing from each other only in their intercept values. A scan rate value of 37 ms was appropriate for all conditions, PM and SM, positive and negative probe, and low versus high interference in SM. For this reason the second major term is simply added to the PM equation to produce the appropriate equation for SM performance. In the instance of PM, where the retrieval process is not required, this term has a value of zero. The fact that the scan rate or slope term is absent from the second major retrieval term has significant implications for the nature of the retrieval act, for it reflects the finding that it takes no longer to retrieve a memory set of four words than a set of two words. This finding led to the conclusion that the set is retrieved as a whole from SM, and not item by item. Once retrieved, the scanning decision making activity proceeds just as in the PM situation. It

is important to note that the second major term is simply added to the first major term to predict SM reaction time.

Experiments 2 and 3: Age, scan rate and retrieval time. This section reports the results of two experiment, one comparing nursery school children (4.5 years) with college students (20 years); the other, college students with an elderly population (70.5 years). The procedures and the materials used in the two experiments differed so direct numerical comparisons between the two studies are not justified but it is possible to make comparisons via their relationship to their control groups.

Procedures. For the child experiment the materials were unrelated words drawn from a vocabulary list for primary grade students; the distractor task was to track a random series of a high or low-tones at one tone per second for 12 sec. The memory sets were 2 or 4 unrelated words. We used a fixed set procedure in which the memory set was first learned by the subject, used for 8 trials (4 positive and 4 negative), followed by a shift to a new set and so on. The materials were presented auditorily by a cassette recorder. Each subject served in both the PM and SM condition. There were 24 subjects in each group.

For the elderly and its control, the materials were presented on a computer screen with the time relationships for SM and PM being the same as in the child experiment. The distractor task was the presentation of 6 two-digit numbers at a 2 sec rate: Subjects named the number aloud and identified it as odd or even. The memory set of 2 or 4 words were drawn randomly by the computer from a pool of 420 words selected from Battig and Montague (1969). They were replaced after each trial. A varied set procedure was used. In both experiments the subjects responded "yes" or "no" by pressing a button with the left or right hand. Time was measured in msc units.

Results. The result of both experiments are presented in Table 1 . The first column identifies the groups, the second the probe type, the third the collapsed mean in msc for the 2 and 4 item sets in PM and the fourth the same measure in SM. The next 2 columns present the scan rate for the combined positive and negative conditions for PM and SM. The final column is the Retrieval Time measure--the difference between the mean PM score and SM score.

Table of Results of Both Experiments

Group Age	Probe	PM	SM	Scan		Retrieval
4.5 yrs	+	992	958	PM	SM	time
	-	1021	1124	24	19	135
20 yrs	+	280	335			
	-	300	363	25	23	59
70 yrs	+	926	1067			
	-	936	1422	22	61	584
20 yrs	+	580	797			
	-	587	777	31	64	204

Discussion

Comparative speed. The statistical analysis showed a main effect of Age in both experiments and for the 70 year group there was also an Age x Memory type interaction. The reaction times for the children for all the data cells were roughly three times those of their control group in both SM and PM. The elderly, however, were about 60 percent slower than their controls in PM but this figure increased to about 95 percent in SM. These results support the position of particular loss in SM with increasing age (Craik, 1977).

The comparative scan rates for all groups are remarkably similar in the PM condition and neither group shows a significant interaction between Memory type and scan rate. In the child experiment there is no reason to suspect that there is a difference between Memory type and scan rate. In the elderly experiment both groups have a scan rate in SM which is about double the PM value. Even though no memory type x set size interaction was found, a difference of this magnitude has not occurred in previous work using this paradigm. However, there is no reason to believe that the elderly group differs from its control with respect to scan rate in either PM or SM.

An important conclusion to be drawn from the research with the elderly is that they differ very little from the controls in scan rate in either memory condition but differ considerably in retrieval time. The elderly also show a large increment in error rate to the positive probe in the SM condition, it being about triple that of their control.

Retrieval time and PM performance. The equation derived from the Wickens et al. (1981) experiment contained two major terms which did not interact and in that sense are independent of each other. In several experiments in which a within-subject design was used with respect to memory type, correlations were computed between PM performance and retrieval time measure was the difference between the PM measure and the same measure for the SM condition. One set of correlations was obtained in the above experiment with children and the other a college student population in a study by Wickens, Moody & Vidulich (1985). Correlations were computed for each population separately, and for both the positive and negative probe. The correlations obtained were all negative and most of them fell between -.25 and -.30, and most did not differ significantly from the zero value. Negative correlations of these values can most parsimoniously be accounted for by the concept of regression toward the mean. Subjects who are exceptionally high or low in one situation would be expected to regress toward the mean on the other test, the amount depending on the reliability of the meausre. Since C. Wickens et al.(1985) found reliability correlations for the PM Sternberg be fairly high (upper .70s) one seems justified in assuming that our correlational data, gathered from fairly homogeneous populations imply that there is little relationshp between the overall score in PM and the important psychological process of retrieving from SM or inactive memory.

General Summary

In the introduction a mention was made about the potential applicability of the research paradigm to the military situation, and I wish to become

more specific at this point and suggest that this paradigm be added to such a test battery as represented by the information processing battery of Wickens, Braune, Stoke and Strager (1985), and for the following reasons.

1. The unique contribution of two-memory paradigm is its capacity to measure the time required to retrieve information from secondary memory. As the correlational data in the last section above indicates, this psychological process does not seem to correlate to any high degree with the PM Sternberg. Thus its use should add to the type of information obtained from the battery.

2. Age does not seem to affect the scanning process to any marked degree but it has considerable impact on retrieval time, which in turn must be involved in many skilled performances.

3. The relative ease with which the paradigm can be instrumented and modified to fit into any program which currently includes the SM Sternberg.

References

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology Monographs, 80, (31, Pt. 2).

Coyne, A. C., Allen, P. A., & Wickens, D. D. (In press). The influence of adult age on primary and secondary memory. Psychology and Aging.

Craik, F. I. M. (1977). Age differences in human memory. In J. E. Birren and K. W. Shaie (Eds.), Handbook of psychology on aging. New York: Van Nostrand Reinhold.

Donder, F. C. (1969). Over de Snelheid van psychische processen. Onderzoekingen godaan in het Physiologish Laboratorium der Utrechtsche Hoogeschool, 1868-1869 (Tweede Roeks, II, 92-120). In W. G. Koster (Ed. and Trans.), Acta Psychologica, 30, 412-431.

Gates, A. I. (1935). A reading vocabulary for the primary grade revised and enlarged. New York: Teachers College, Columbia University.

James, W. (1890). Principles of psychology (Vol. 1). New York: Holt, Rinehart & Winston.

Karsko, J. K. (1979). A comparison of retrieval time and scan rate of preschool children and adults. Dissertation, Ohio State University.

Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction-time processes. American Scientist, 57, 421-457.

Wickens, C. D., Braune, R., Stokes, A., & Strager, D. (1985). Individual differences and age-related changes: Refinement and elaboration of an information processing performance battery with aviation relevant task structures. Final Technical Report EPL-85-1/NAMRL-85-1.

Wickens, D. D., Moody, M. J., & Dow, R. (1981). The nature and timing of the retrieval process and of interference efects. Journal of Experimental Psychology: General, 110, 1-20.

Wickens, D. D., Moody, M. J., & Vidulich, M. (1985). Retrieval time as a function of memory set size, type of probes, and interference in recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 154-164.

Effect of Vertical Heave Motion on Cognitive Performance

Kim Wilson and Jay Pollack
Naval Biodynamics Laboratory

Abstract

Previous research examining cognitive performance during motion sickness-inducing stimulation (e.g., vertical heave) has been equivocal. The purpose of this research is to determine cognitive performance effects of ship motion using the Naval Biodynamics Laboratory's Ship Motion Simulator (SMS). Six human research volunteers participated in a single-subject repeated measures design. This design was employed to eliminate potential confounding due to group analyses which may mask individual behavior during motion conditions. While underg. .g large amplitude heave motion, subject performance on a battery of performance tests was measured. This report examines the effects of motion on a test of complex problem solving. Each subject's performance during motion conditions was compared with the baseline performance established prior to the beginning of the motion sessions and measured periodically throughout the 15 weeks of the study. Analyses of the data supported the hypothesis that motion effects on performance are idiosyncratic across subjects, but consistent within subjects. Performance decrements did not occur for all heave conditions studied.

Background

Recent expansion of Soviet naval capabilities has required the U.S. Navy to improve its ability to operate effectively in areas of northern latitude. The seaways in these operating areas are frequently rough, leading to large amplitude, low frequency ship motions. In 1975, VADM R. E. Adamson reported on the effects of ship motion on human performance and stated that the operational performance of the fleet during inclement weather was significantly affected by ship motion. Ship motion caused performance decrements can be attributed to several sources including biomechanical interference, reduced cognitive performance, kinetosis, and overall poor habitiability leading to acute loss of health and well being. The Naval Biodynamics Laboratory (NBDL) is currently exploring the effects of ship motion on these factors and proposing ameliorative procedures to improve ship all-weather fighting capability. In this report, we describe the effects of heave motion on cognitive performance and kinetosis on a small group of navy enlisted personnel. In future reports, we will present data obtained on other performance tests (NBDL Performance Assesment Battery (PAB86)) administered in conjunction with the heave conditions reported here.

The increasing reliance on automated weapon systems onboard Navy combatants has resulted in increased cognitive demands upon weapon system operators. Nearly all current theories of cognitive functioning assume that there are limited attentional resources that can be used for performing any particular task. If several tasks compete for the same resources, one or more of the tasks will be performed less well (Wickens, 1980). We propose here to generalize this theory by viewing the subjects response to ship motion as a task which competes for resources. Consequently, the situation where an individual is performing a cognitive task during conditions of ship motion is similar to situations where the individual is performing several tasks at once. As individuals experience motion, they begin to monitor both the external environment and their internal responses to the environment. This increased monitoring will compete for cognitive resources and may result in decrements in task performance in much the same way that performance in multiple task situations is affected. Whether performance decrements are actually observed depends on (a) whether a particular motion condition results in increased external and internal monitoring and (b) the amount of cognitive resources that an individual must devote to successfully perform the task.

The results of prior studies examining the effects of low frequency motion on performance have been equivocal. Based upon the theory of low frequency motion effects on cognitive performance described above, there are several possible reasons as to why past studies have not established any clearcut relationship between cognitive performance and low frequency motion. First, there are individual differences in psychophysiological responses to low frequency motion. It is reasonable to assume that the amount of cognitive resources allocated to monitoring internal and external states will be somewhat dependent upon the individual's psychophysiological responses to motion and also be dependent upon the individual's idiosyncratic style of dealing with psychophysiological stress. Second, there are individual differences in the amount of cognitive resources an individual must devote to perform a task successfully. If an individual has to allocate virtually all of his cognitive resources to the task, any reallocation of cognitive resources to external or internal monitoring will lead to a performance decrement, whereas individuals who can successfully perform the task with only minimal demands on their cognitive resources will be less likely to experience resource competition and its disrupting effects on performance.

Consequently, group analyses of the effects of motion on performance may mask the effects of motion. Some individuals in the group may be expected to exhibit performance decrements while other individuals will not exhibit performance decrements. The likely result of this is that there will be a moderate decrement in the group performance, but this decrement will not necessarily be striking in size. Given the potential for highly idiosyncratic effects of motion on performance, a more appropriate and more sensitive method for studying the effects of motion on performance is to use a single-subject design. The research reported here examines the effects of heave motion on cognitive performance for six individuals for whom baseline performance data were collected throughout a three-month period for comparison with data collected during heave motion.

METHOD

Subjects. The subjects were six males, ages 19 to 28. They were active duty enlisted Navy personnel who requested assignment to the Naval Biodynamics Laboratory (NBDL) subject pool. Complete

data for only five subjects is reported here as one of the subjects dropped out because of a particularly severe reaction to motion.

Apparatus. Vertical heave stimulation was obtained using the NBDL Ship Motion Simulator (SMS). This device is a 2.5 cubic meter room capable of simultaneous heave, pitch and roll. The prime mover is a large amplitude, single-jack, vertically acting electrohydraulic oscillator acting in the range from 0 to 5 Hz under servovalve or computer control. Roll and pitch are provided by a yoke-and-trunion system attached to the undercarriage of the moving cab. The cab is windowless, and was configured with seats and PAB86 administration systems.

The battery of performance tests (PAB86) was developed to run on APPLE II computers with 13 inch NEC visual display units (VDU). The subjects sat at a work station approximately 18" from the VDU.

The SMS was equipped with three testing stations. The stations were separated by a vinyl sheet that made it impossible for the subjects to see one another or see any other VDU besides their own. The subject's used an APPLE KEYPAD to make their responses.

Description of Task. Subjects were administered the PAB86 during both baseline and motion testing sessions. The PAB86 consists of 5 seperate performance tests separated by either a vigilance test (unscored) or a motion sickness questionnaire. Only the complex problem solving (CPS) test results are reported here. The CPS test consists of a matrix of 12 cells (3 rows X 4 columns) displayed on the VDU. Each cell was approximately one inch square. The cells were numbered consecutively from 1 to 12, with cell number one the upper leftmost corner square and cell number twelve the lower rightmost square. The cell number was displayed in the middle of the cell. The cells always appeared in the proper numbered sequence on the screen. For each test administration, the computer generated a target sequence (using the numbers 1-12) to be discovered by the subject in solving the problem. Pressing a correspondingly numbered key on the keypad would result in covering the number displayed in that cell. If the number entered by the subject was not in the proper sequence relative to the other numbers already entered, the cell numbers that were out of sequence would uncover. This is illustrated in Table 1. The subject's task was to cover all of the cells as rapidly as possible while making as few entries as possible.

Design. Six subjects were randomly assigned to two groups. One group received all testing in the morning and the other group received testing in the afternoon. Acquisition training took place over a three-week period prior to any motion runs. Each subject received a minimum of 10 acquisition trials during this period.

Data collection took place over a 12-week period. An individual data collection session lasted approximately two hours. During the two-hour period, the subjects were administered the PAB86. The battery took approximately 20-25 minutes to complete and was administered three times during the data collection session. During the testing session, when the subjects were not working on the performance test battery, they performed a simple vigilance task. The subjects performed the PAB86 in the no motion (static) condition on Monday, Tuesday and Friday, and performed the PAB86 under motion conditions on Wednesday and Thursday. There were five different heave only motion conditions and five different roll only motion conditons (the inclusion of the roll conditions was to provide additional baseline comparison data). The heave conditions were administered in a different random order for the two groups.

Motion Conditions. The motion conditions examined in this study were all simple sinusoidal heave motions. These simple heave motions can be described in terms of frequency and RMS acceleration. Table 2 lists the five heave conditions. The heave conditions are ordered with respect to increasing expected motion sickness incidence.

RESULTS

Given this model regarding the idiosyncratic effects of motion on cognitive performance, it was necessary to use analytical procedures that are appropriate for single-subject studies. The general plan for analysis was to contrast each subject's performance on the CPS task during each motion condition with the subject's baseline performance. The dependent measure was time to complete the task. This analysis was carried out for each subject separately.

A single value estimate of baseline performance was calculated by collapsing across all static test trials administered on the Tuesday and Friday testing sessions. The data collected on the Monday sessions were not included in this calculation because performance on these sessions was expected to be depressed due to the weekend layoff. We were concerned that the baseline performance estimate would possibly be an underestimate of the subject's actual baseline performance. This would result in underestimating the effects of motion because motivation would likely be higher during the actual motion sessions than on the static data collection sessions. Consequently, an additional estimate of baseline performance was made by collecting CPS data during roll-only motion runs. Pure roll motion has been demonstrated to be generally unprovocative of motion sickness and was not thought likely to cause performance decrements on the CPS task. Again, a single value estimate of performance was computed by collapsing across all of the test trials adminstered during the roll conditions.

A single value estimate of motion performance was made for each subject by collapsing across all of the motion data points. Table 3 summarizes both baseline performance estimates and the mean motion performance for each subject. Since these data indicate that both baseline estimates yield approximately the same values for all of the subjects, all analyses were conducted using static performance as the baseline comparison data. A simple t-test was conducted comparing motion performance with baseline performance for each subject, using the logarithms of the time to complete the task as the dependent variable. Transformation of the data was necessary because time data are not normally distributed and a single large outlier would skew the analyses. This statistical model was judged to be appropriate because, although all of the data comes from the same individual, the

error associated with each measurement is due to exogenous factors and, hence, the errors are uncorrelated. Because the assumption of homogeneity of variance becomes very important when dealing with N's of widely different sizes, the test statistics were evaluated using the corrected degrees of freedom.

For the purposes of this presentation, a more fine-grained analysis was also conducted. Each subject's performance during each motion condition was estimated by collapsing across all of the data points available for that condition. If the subject did not abort the run, there were a maximum of six data points for each motion condition. When a subject aborted a motion run and failed to complete the entire session because he could no longer perform the required task, we made the assumption that performance was completely degraded and time to complete the task was infinity. Figure 1 summarizes each subject's mean performance during each of the heave conditions (instances where a subject's time to completion was infinity are represented by cross-hatching). The solid horizontal line on each graph represents that subject's static baseline performance and the dotted horizontal line represents that subject's roll-only baseline performance.

Tests of statistical significance of the difference between performance during each motion condition and the baseline conditions were made by using a simple t-test on the logarithms of the time to complete the task. Table 4 summarizes the results of these tests for each subject. The express purpose of the research was to determine whether there are performance decrements associated with ship motion. Consequently, one-tailed significance tests are reported.

An additional analysis was conducted using more conservative, non-parametric techniques. A z-score for each motion data point was computed based on the distribution of the baseline performance scores. A one-sided 95% confidence limit was calculated for the distribution of baseline performance scores and each motion z-score was compared to that confidence limit. Table 5 presents the percentage of data points for each motion condition that exceeded that confidence limit. A chi-square was calculated for each motion condition by summing the squared z-scores. These chi-squares are reported in parantheses in Table 5. The chi-square value was evaluated using six degrees of freedom.

Motion Sickness. Two of the five subjects vomited during a motion session. Subject #160 vomited during condition 5 and subject #165 vomited during motion conditons 4,5,6. The subjects were allowed to terminate a motion session if they became so ill that they did not feel the could continue to function any longer in that environment. The determination to terminate a motion session was not necessarily related to vomiting episodes. That is, a subject did not have to vomit to terminate a motion session, nor did vomiting necessarily mean a session was terminated. There were four subject terminated motion sessions during the course of the study. Subjects #165 and #171 aborted during condition 3 and subject #160 aborted conditions 3 and 5. In general, both the episodes of frank motion sickness and subject initiated premature termination of motion sessions occurred in concert with the expected increased severity of motion as expected.

DISCUSSION .

The static test performance flucuations present in the current study reflect the performance flucuations one observes in actual field operations where an individual's particular problem solving strategy does not always provide the most efficient solution to all problems and level of motivation changes frequently. It was our plan to employ a test which would respond the same way as field operations to daily changes in motivation and strategy. Other researchers concerned with evaluating environmental effects on performance have raised arguments for employing tests that yield very stable performance across days. However, tests that yield extremely stable performance from day to day may possibly become highly automatized. As a process becomes automatized, it requires fewer and fewer cognitive resources to perform and consequently, the test used to measure that process will become less sensitive to environmental effects that arise from resource competition. This pitfall was avoided with the CPS test.

A model based on an information processing analysis of motion effects was proposed that predicted idiosyncratic effects of motion on cognitive performance even in the absence of overt symptoms of motion sickness (e.g., nausea and vomiting). A particular motion condition will affect an individual's performance proportional to the resources diverted to monitoring the environment (i.e., ship motion). There will be individual differences i the amount of resources an individual diverts to monitoring the environment and, consequently, the degree to which performance decrements occur. This model is supported by the results of this study. The performance of subjects #165 and #171 was consistently affected by motion. It is especially important to note that there were performance decrements even during those motion conditions where these individuals were able to remain in the environment and continue to perform for the entire two-hour session. On the other hand, subjects #170 and 178 exhibited minimal, if any, performance decrements associated with motion. Subject #178 did exhibit poorer performance during the most severe motion condition. Subject #160 presents a less clearcut picture. On two occasions he aborted the run because he was no longer able to tolerate the motion. On these occasions, the motion environment completely prevented performance. However, on the occasions when #160 was able to complete the session, there is no evidence that motion had any affect on performance. The model proposed clearly predicts that motion will have a differential effect on individual's performance. This differential effect was clearly present in this study. The model also specifies that performance effects are due to a diversion of cognitive resources. The study reported here does not directly address this issue. Further research at this laboratory exploring the locus of performance effects as predicted by this model is underway.

When motion-induced illness affects individuals to the extent that they are no longer able to function, performance may be said to be terminally degraded. However, this study clearly demonstrated that catastrophic failure due to motion sickness, although possible, is not the only way

that motion can affect cognitive performance. It is our opinion that this study, as is the case in any laboratory study of this nature, will underestimate the severity of performance decrements due to motion in the general population. This is because the voluntary nature of subject participation preselects for motion sickness resistant individuals. For example, one subject originally included in this study, dropped out completely because of his particularly strong adverse response to motion and no data could be collected on his performance.

There are several factors in the model that help determine whether or not any particular individual will suffer from performance decrements when working in a motion environment. First, the individual's style of coping with stressful external environments and uncomfortable (even painful) internal states will determine how successful the individual is in performing in a motion environment. Some individuals will be more likely to quickly divert cognitive resources from task performance to external and internal monitoring than will other subjects. Second, the specific nature of the cognitive task to be performed in the motion environment will play a central role in determining whether performance deteriorates in motion environments. Wickens (1980) has pointed out that performing tasks simultaneously will result in poor performance only when the tasks use the same specific cognitive processes. This leads to the conjecture that tasks that require similar cognitive processes as those used in monitoring the environment and monitoring internal states will be those tasks most likely to be disrupted by motion.

Human performance on complex cognitive tasks varies from day to day. There are numerous factors operating in concert to produce this variation, one factor of considerable importance is the extent to which the proper cognitive processes are brought to bear on the task. When other events compete for those same resources, task performance suffers. The question of operational relevance is whether the competition for cognitive resources resulting from ship motion environments produces performance decrements greater than those that might arise on a routine, daily basis. The data reported here clearly indicate that such is the case.

REFERENCE

Wickens, C. (1980). The structure of attentional resources. In R. Nickerson and R. Pew (Eds.), Attention and performance VIII. New Jersey: Lawrence Erlbaum.

TABLE 1

Computer generated target sequence : 13542...

ENTRY	Result on Screen
1	1 covered
3	1,3 covered
4	1,3,4 covered
2	1,3,4,2 covered
5	1,3,5 covered and 4,2 uncov.

TABLE 2

Condition	1	2	3	4	5
Frequency (Hz)	.10	.35	.35	.15	.25
Acceleration (g)	.05	.10	.20	.10	.20

TABLE 3

SUBJECT	ROLL			STATIC			MOTION			T value
	X	S.D.	n	X	S.D.	n	X	S.D.	n	
165	94	25.9	63	95	24.2	27	114	30.4	29	3.39*
178	70	19.0	51	70	17.7	21	79	31.4	25	1.16
170	99	30.7	58	91	25.1	30	106	31.3	33	0.96
171	121	39.0	53	112	30.1	30	164	90.0	22	2.66*
160	156	42.3	43	161	42.8	27	148	43.4	16	-0.67

* p<.05

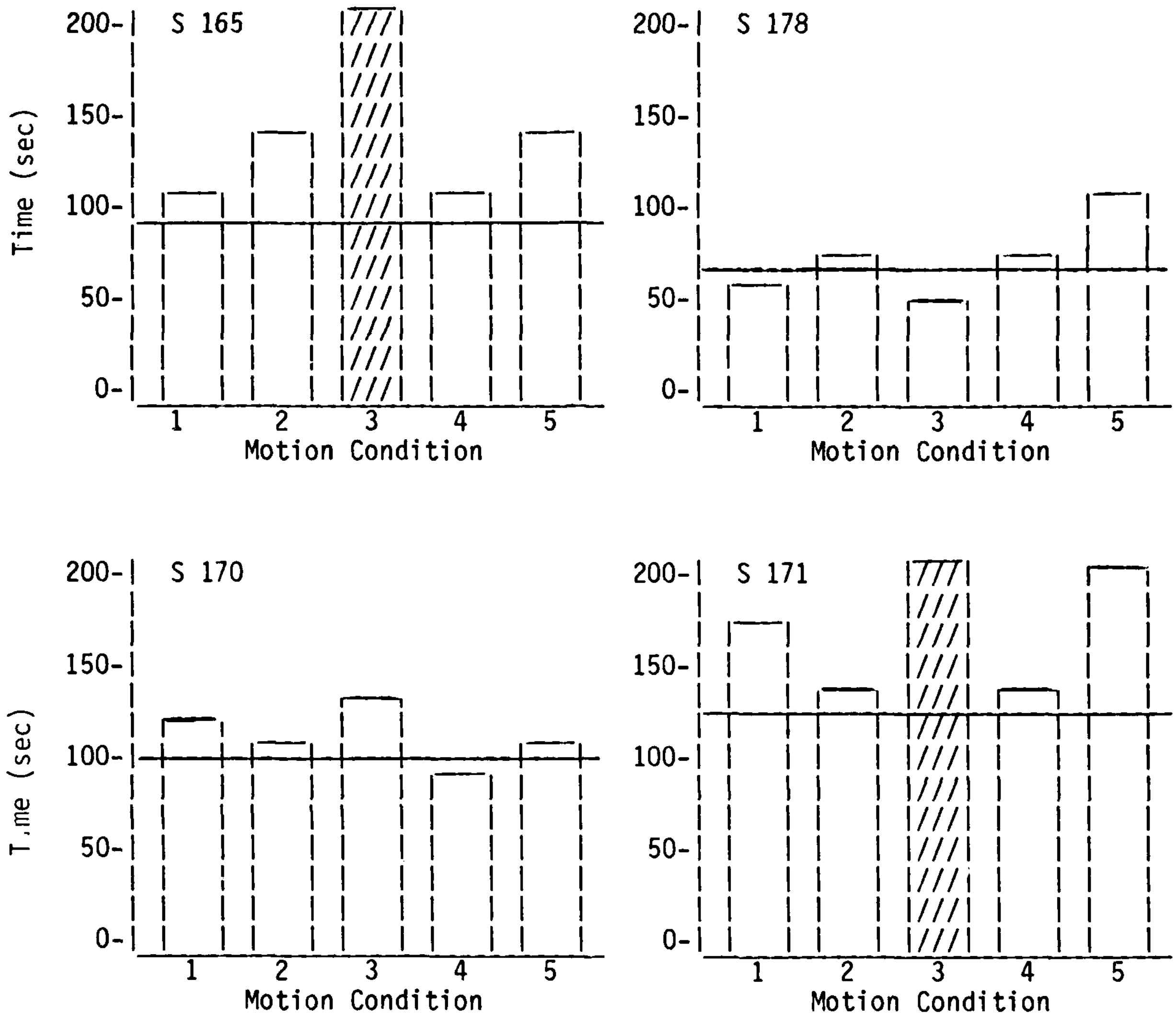
TABLE 4

SUBJECT	Condition 1	2	3	4	5
165	2.92*	3.61*	---	.82	3.97*
178	.60	.36	-1.22	1.05	2.33*
170	1.16	-.09	1.82	-.17	.54
171	1.71	1.30	---	1.39	1.54
160	-.47	1.32	---	.87	---

* p<.05 --- subject abort, infinite completion time

TABLE 5

Subject	Condition 1	2	3	4	5
165	0 (3.2)	50 (21.9*)	---	17 (.70)	50 (21.0*)
178	17 (7.7)	17 (10.8)	0 (1.1)	0 (4.1)	30 (45.8*)
170	17 (10.9)	17 (9.2)	17 (4.2)	17 (8.1)	0 (1.9)
171	50 (26.3*)	17 (6.3)	---	0 (1.5)	33 (104.3*)
160	0 (2.0)	0 (2.3)	---	20 (7.6)	---



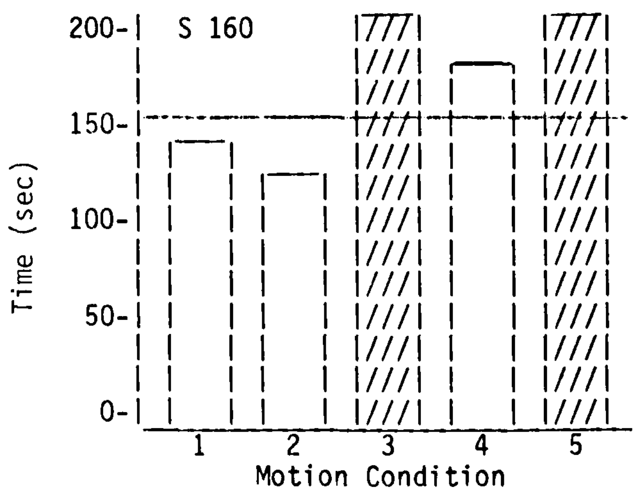


Figure 1. Average completion time for five motion conditions (see text).

Differential effects of articulatory suppression and randomization on the performance of different aspects of a computer game.

David B. Porter
Oxford University

A rapidly-paced, visuo-manual computer game composed of substantively different subtasks (one uncertain and the other complex) was combined with metronome-paced auditory-verbal side tasks involving articulatory suppression in either a fixed or subject-generated "random" sequence. After practicing the game, each of 12 male and 12 female paid subjects performed all combinations of side tasks and game prioritizations. Articulatory suppression in a fixed sequence impaired performance of the uncertain task but had no effect on the complex task. In contrast, the requirement to produce responses in a "random" order initially caused decrements in both subtasks. After thirty minutes practice, randomization continued to significantly impair the uncertain task but had no effect on the complex task. The semantic relevance of the articulated material had little influence on these effects.

Adequate performance is the preeminent goal of the human information processing system. The achievement of this goal is comprised of three functionally distinct conceptual stages (i.e., translation of external processes into internal symbols, combination and transformation of these symbols, and retranslation of symbols into processes (Craik, 1943)). The second of these (the intermediate processes) is both the most interesting and the least tractable. The flexibility and isotropic nature of central systems pose difficulties for traditional empirical methods. In this context the importance of theoretical distinctions is apparent. Distinctions provide "epistemological joints" which serve as references for systematic inquiry.

After presenting a critical review, Neumann (1984) outlines a theory which defines automaticity as a mode of parameter specification:

> "A process is automatic, if all its parameters are specified by a skill (a procedure stored in long term memory) in conjunction with environmental information. If these two sources of constraint cannot specify all parameters, further constraints must be provided by attentional mechanisms."(p.293)

Neumann uses "attentional mechanisms" in much the same way others employ the term "resources". These cognitive entities are of limited capacity, agnostic, under direct intentional control and provide alternative means to perform intermediate transformational functions.

Another important distinction concerns the differentiation of "resources" themselves. Early theorists assumed that processing resources were undifferentiated and that the amount of resource required by a particular task was reflected by the task's "difficulty". Later theorists have moved toward "multiple resource" formulations, which specify different types of resources. The similarity between two tasks is seen to be more predictive of dual task interference than difficulty. "Working memory" is a concept which incorporates both these resource notions (Baddeley & Hitch, 1974). Working memory is an intermediate processing system composed of both general-purpose resources (i.e., the central executive) and separate, more specific resource mechanisms (i.e., the articulatory loop and the visuo-spatial scratch pad). Processes within resource mechanisms are still controlled by the central executive but these mechanisms facilitate processes and augment capacity for certain types of tasks. The articulatory loop has been studied extensively and is closely related to verbalization. A subsidiary question concerns the extent to which the semantic contents of verbal material being processed are insulated from the semantic content of other material being processed.

If interference caused by the similarity of peripheral processes is controlled, these distinctions allow several predictions concerning interference patterns during dual task performance. In order of increasing generality these predictions are:
1) To the extent information being processed is unencapsulated, greater interference

will be caused by side tasks involving semantically significant material.
2) To the extent a side task can be confined to a specific resource mechanism (i.e., the articulatory loop), it will only interfere with those tasks which are facilitated by the employment of that mechanism.
3) To the extent a side task involves general processing resources, it will cause more general interference with performance.
4) To the extent a particular task involves constant rather than variable or indeterminant structure-relations, its performance will become automated and less subject to interference from side tasks involving even general processing resources.

PROCEDURES

Twelve male and twelve female subjects between 18 and 38 years of age from the Oxford Subject Panel individually participated in this two-hour experiment. A 48 K Sinclair Spectrum microcomputer and 14 inch color monitor were utilized.

Subjects completed 150 trials of a four-choice reaction task. The index and middle fingers of both hands were overlapped on the keyboard to produce a direct spatial mapping for responses. Subjects next received training in controlling the movements of a blue whale (the protagonist of the game). Although the controls were the same as in the reaction task, subjects were required to hold the key down until the whale responded to compensate for a variable control lag. Subjects received as many practice trials as necessary to reach the criterion (10 laps around an iceberg in less than 70 seconds).

Subjects completed 6 practice trials of the Save The Whale computer game. They were then taught to perform the verbal side tasks and completed an additional 18 trials combining different game prioritizations and side tasks. Interpreting the experimental results depends on understanding the computer game itself; this will be described in some detail.

The monitor screen displayed a 21 X 31 cyan matrix containing four clusters of white icebergs as shown in Figure 1. Subjects controlled the direction of movement of a blue whale (initialized at screen center for each trial) by using keyboard controls. The whale had two substantively different tasks: eating plankton and wrecking kayaks. The differences between these tasks are important and more fully described in the preceding paper (Porter, 1986). In order to score points by "eating plankton" subjects had to guide their whale through a small flashing mass of plankton. The plankton followed an irregular but relatively continuous psuedo-random diagonal path. This task was simple but highly unpredictable. Kayaks were launched from one of five unmarked locations at the screen border and moved one space toward the whale on each successive cycle. If they reached the whale, the subject lost points, but if the kayaks "crashed" into an iceberg instead, subjects gained points.

A trial lasted 2 minutes 15 seconds and consisted of 218 "cycles" during which each of the characters could move one space. The two tasks were combined in three ways. In plankton priority trials, plankton was worth 100 points and crashing or being harpooned by kayaks was worth plus or minus 10 points. In the kayak priority trials the points were reversed. In equal priority trials all contingencies were worth 50 points. The priority was explained by instructions displayed on the screen before each trial. Half the subjects began with a kayak priority trial, followed by equal priority and then plankton priority. The other subjects accomplished the trial priorities in the reverse order.

Having completed 6 practice trials (2 of each priority), subjects received side task instruction. Because the game was visuo-manual, audio-vocal side tasks were selected to minimize interference caused by peripheral similarity. Two different verbal side tasks with two different sets of contents were employed. Each set of contents contained four single syllable words. "One", "two", "three", and "four" comprised the control set and "left", "right", "up", and "down" made up the semantically-relevant set. The four directions corresponded to the position and function of the four keys subjects used to control the whale's movement. Two different verbal side tasks required subjects to produce responses from one of the sets in time to a mechanical metronome at a pace of one response every 1.5 seconds.

The difference in the two side tasks was the order in which subjects were to produce the responses. In the simple articulatory suppression condition, subjects produced responses in a fixed order (e.g., "1,2,3,4,1,2..."). It was assumed subjects could relegate this task to the articulatory loop with only minimal involvement of general processing resources. In contrast, the other side task placed greater demands on general resources. For this task (Baddeley, 1966), subjects were required to produce responses in a "random" order. Before attempting this side task, the concept of randomness was briefly discussed. Subjects were told that a randomly generated sequence should have approximately equal frequencies of each of the possible responses, that there should also be an equal frequency of the 16 possible "digrams", and that trends should be counterbalanced. Subjects provided 100 responses in time to the metronome, first using numbers then directions. Their verbal responses were immediately put into the computer by the experimenter. At the completion of each run, subjects were debriefed on their performance on each of the criteria. All subjects were familiar with the concept and simulation of randomness.

After a short break, each subject completed 18 successive trials (6 of each of the priorities in rotation). For each of the three priorities, two of the six trials involved no verbal side task, two involved simple articulatory suppression and two involved randomized articulatory suppression. One of two trials of each priority (with each of the side tasks) required "numeric" responses and the other required "directional" responses. Thus, each subject completed each of the possible combinations of priority, side task type, and side task contents. Order of presentation was counterbalanced across the 24 subjects. At the end of each trial, subjects recorded their scores on both subtasks and total points from the monitor display and were told their "randomness" score by the experimenter. After the final trial, subjects were debriefed and paid. Fatigue did not appear to be a major factor and many subjects commented that they had enjoyed the game despite the side tasks.

RESULTS

Means and standard deviations reflect general performance levels. The most important data, however, are those showing the effects of the different conditions. Raw measures were "standardized" by subject across the 24 trials. Multiple regression analyses with the appropriate "degrees of freedom adjustments for repeated measures" (Cohen & Cohen, 1983) were employed to analyze the within-subject variance. Significant interactions as well as main effects are depicted by plotting mean standard scores for each criteria.

Subjects were capable of performing the subtasks. Overall, subjects changed the direction of the whale's travel 64.6 (SD=12.7) times (on 30% of the 218 cycles or about every 2 seconds). Their whales consumed an average of 11.1 (SD=10.8) tons of plankton and wrecked 9.4 (SD=3.9) kayaks per trial. When plankton was afforded priority, the mean tons of plankton eaten was 21.0 (SD=7.1). Similarly, when kayaks were top priority, subjects were able to wreck 12.7 (SD=2.3) kayaks. In other words, on average, subjects were capable of "scoring" on plankton about 10% of the time or wrecking about two thirds of the 19 kayaks launched each trial.

The following multiple regression equations show the main effects (as partial regression coefficients) of the priority instructions (PRI), pracitice (RUN), simple articulatory suppression (ASUP), and randomized articulatory suppression (RDZ) on the two standardized criterion scores: plankton eaten (PE) and kayaks crashed (KD). Variables reflecting all possible two-way interactions were examined and the two found to be significant (the interaction of practice and priority (PXR) for the plankton task and the interaction of randomized articulatory suppression and practice (RXR) for the kayak task) were included in the respective equations. Corresponding t values are shown below each of the predictors. The R', degrees of freedom, and level of significance for each equation is also listed.

PE =	7.80 PRI -	.48 RUN -	2.02 ASUP -	4.36 RDZ +	.52 PXR + 34.36
t(df=324)	6.72**	-1.12	-2.93**	-6.32**	2.60**

R' = .76 F(90,324) = 11.46 p<.01

$$KD = -9.50\ PRI + .43\ RUN - .42\ ASUP - 6.97\ RDZ + .94\ RXR + 67.30$$
$$t(df{=}324){-}24.36^{**} \quad 1.87 \quad -.53 \quad -2.97^{**} \quad 2.35^{*}$$
$$R' = .65 \quad F(90,324) = 6.72 \quad p<.01$$

Articulatory contents (i.e., verbalizing directions instead of numbers) did not significantly predict performannce in either task ($t(df,306)=-.70$ for the plankton task and $t(df,306)=+.61$ for the kayak task). This variable was excluded from the regression equations to preserve the significance of the other independent variables and interactions.

Of the within subject variance in the plankton task, 76% was explained by the five predictors. Priority instructions (coded 1 for kayaks, 2 for equal and 3 for plankton priority) strongly influenced performance. The significant positive interaction between priority and practice (PXR, the product of priority and practice) suggests that performance on the plankton task improved when plankton was priority but not when kayaks were priority. Both simple and randomized articulatory suppression (each variable coded either 0 or 1) significantly interfered with performance but randomized suppression caused the greater disruption. These deleterious effects were not significantly ameliorated with practice.

Priority was also a significant predictor of performance on the kayak task, but did not interact with practice. This suggests that the slight improvement in performance with practice was shared more equally across all priority conditions. Simple articulatory suppression did not significantly influence performance on the kayak task. In contrast, randomized articulatory suppression interfered strongly with performance initially, but the positive interaction between randomization and practice (RXR, the product of the two) suggests this effect was diminished significantly with practice.

The multiplicity of effects is somewhat difficult to vizualize. Plotting the mean standardized scores under different conditions when each task was to be afforded top priority allows direct graphic comparison of the effects. These scores are plotted in Figure 2 for each task in the three conditions for the first half of the experimental trials (broken lines) and then the second half (solid lines).

The additivity of effects of the two experimental conditions (i.e., articulation and randomization) on the plankton task is shown. Although there is improvement with practice, the additive effects remain. In contrast, the difference in the effects of the two side tasks on performance of the kayak task is evident. In both runs simple articulatory suppression had a slightly facilitory effect on performance. During the first half of the experiment, the interference in kayak task performance appears to be nearly identical to that of the plankton task. The significant interaction between practice and randomization for the kayak task is shown by the abolition of interference during the second half of the experiment.

DISCUSSION

There was little support for the hypothesis that the semantic content of the verbal side task would influence the main task. In this experiment, the type of processing activity was a much more potent predictor of performance than was the semantic content of the activity.

It was assumed that simple articulatory suppression could be relegated to the articulatory loop and thus exert relatively specific interference effects. Earlier findings (Porter, 1986) suggested that the articulatory loop facilitated performance of an uncertain task (i.e., plankton eating) but not a complex one (i.e., kayak wrecking). The results of this experiment support both the assumption and the earlier finding; simple articulatory suppression interfered with performance of the plankton but not the kayak task. The argument that the kayak task is simply impervious to interference is diminished by the nearly equivalent decrement caused to both primary tasks by randomized articulation. These results also provide support for the third hypothesis (i.e., that side tasks involving general processing resources cause more general disruption). To the extent the plankton task reflected a relatively indeterminant relation-structure and the kayak task reflected a more

constant relation-structure, the final hypothesis also receives support. The significant interaction between the effects of practice and randomization suggests subjects were able to "automate" performance of the kayak task, thus insulating it from the interference effects of even randomization. This was not possible for the interference caused by either of the side tasks to the plankton task.

Beyond these specific conclusions, these results also suggest claims that theoretical distinctions based on laboratory-type experiments have little relevance to the performance of meaningful tasks, are somewhat overstated. Likewise the implicit assumption that psychological experiments must be gruelling experiences in order to yield interpretable results might also warrant reexamination.

REFERENCES

Baddeley,A.D. (1966). Capacity for generating information by randomization. Quarterly Journal of Experimental Psychology, Part 2.

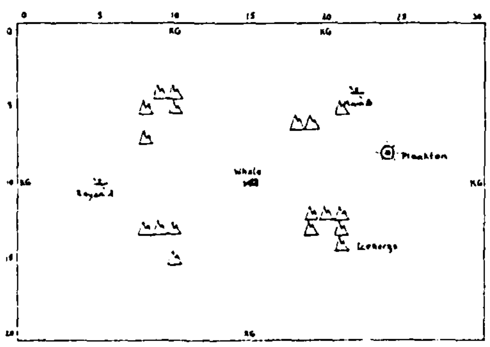
Baddeley,A.D., & Hitch,G. (1974). Working memory. In G.H.Bower (Ed) The Psychology of Learning and Motivation, Vol.8. London: Academic Press.

Cohen,J., & Cohen,P. (1983) Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Hillsdale, N.J.: Erlbaum.

Craik,K.J.W. (1943). The Nature of Explanation. Cambridge: University Press.

Neumann,O. (1984). Automatic processing: A review of recent findings and a plea for an old theory. In W.Prinz & A.F.Sanders (Eds.) Cognition and Motor Processes. Berlin: Springer-Verlag.

Porter,D.B. (1986) Differential effects of memory load on the performance of substantively different aspects of a computer game. Proceedings: Tenth Psychology in the Department of Defence Symposium. Colorado Springs, Co: USAFA.



Monitor Display

Fig 1

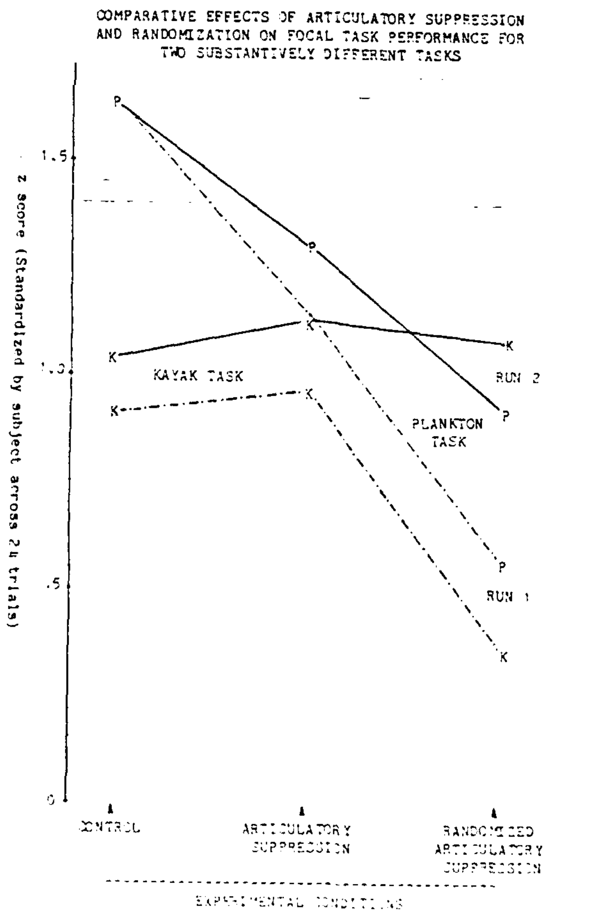


Figure 2.

IMPROVING THE PLACEMENT AND RETENTION OF NAVAL SHIPYARD WORKERS

Herbert George Baker, PhD

Navy Personnel Research and Development Center
San Diego, California 92152-6800

Abstract

This paper reviews current efforts to assist the Naval Sea Systems Command (NAVSEA) in improving its interviewing and classifying procedures that are used when hiring blue collar shipyard employees. NAVPERSRANDCEN is developing: (1) a structured interview protocol for use by the employment interviewers; and (2) a number of components designed to broaden the placement process, including computer-based comprehensive occupational information, audiovisual materials that provide orientation to shipyard work and realistic job previews, and an on-line preference elicitation instrument. These improvements should conduce to better informed employment decisions by both the organization and the individual, resulting in more optimal placement, increased job satisfaction, and decreased personnel turnover in the blue collar trades.

Introduction

To fill a variety of skilled and semi-skilled jobs, the Naval Sea Systems Command (NAVSEA) primarily relies on local hiring at the particular shipyards. A majority of job applicants for NAVSEA trades openings are young persons, 17-22 years of age, with little or no work history and therefore no developed job skills. It is, in this sense, a hiring situation paralleling that of armed services recruiting (Baker, 1985a).

Young job applicants are usually very career naive, and know little about their own abilities, interests, and job preferences: even less about the world of work into which they seek entrance. They are equally unknown to the prospective employer (Baker, 1983).

Furthermore, even though they may have attended job fairs and received information about NAVSEA jobs, the expectations of such applicants are frequently unrealistic. Their expectations require validation or even confrontation. Likewise, the employing organization requires means to refine its expectations of the individual's potential for satisfactory performance and tenure (Baker, 1985b).

Placement interviews, then, must provide the means for such things as applicant assessment, occupational exploration, and refinement of expectations. In short, placement is the end result of a person-job matching (PJM) process of whatever sophistication.

PJM is thus a bridging mechanism, linking individuals with jobs. The PJM process is of immediate criticality to worker placement. However, its more lasting impact is on organizational productivity and individual job satisfaction (Baker, 1983).

Current Methods and Problems

Job fairs and other informational programs serve to acquaint prospective job seekers with opportunities for employment at the shipyard. To assist in qualifying job applicants, an aptitude test provided by the Office of Personnel Management (OPM) is used to assess the cognitive dimension. Eligibility for employment is based on attainment of a passing score on that test. Each shipyard maintains its own roster of eligibles, who are rank ordered according to their OPM test scores. Selection is, thus, a go/no go decision.

Job offers and subsequent classification are accomplished during a placement interview. Typically, that placement interview is conducted by the general foreman who has cognizance over the trades for which hiring is taking place.

The interview itself is loosely structured and of widely varying format. Because of this, the techniques employed do not uniformly elicit the same kinds of personal information from job interviewees. On the other hand, the procedures do not include any means of providing personalized, comprehensive occupational information to the job applicant, nor do they include sophisticated means by which the applicant may indicate occupational preferences. During the placement interview, there is no general orientation to shipyard work and no specific job previews.

Consequently, current procedures do not assist either the interviewee or the interviewer in making wise occupational decisions. The interviewer is forced to rely on inadequate information about the applicant. Likewise, the applicant has no means of weighing personal characteristics against available jobs.

Suboptimal placement is, under these conditions, a genuine risk. Trades openings are apt to be filled with individuals who have perhaps minimal career maturity and who are operating under unvalidated or even erroneous job expectations. Once on the job, disconfirmed expectations and work under unforseen conditions frequently leads to dissatisfaction. Consequences of job dissatisfaction include lowered morale, decreased productivity, and shortened tenure (Baker, 1985a).

There are indications that the procedures used to place Naval shipyard tradesmen are in need of sophistication and refinement. Indications of this are unacceptably high turnover rates among employees. Of particular importance is the high turnover among workers in the skilled trades, the apprentice programs, because of the costly training investments in those workers. Shipyard management desires to refine the placement process in order to more optimally classify new workers and improve worker tenure.

Research Objectives and Approach

NAVPERSRANDCEN has undertaken to design, develop, and assist in the implementation of a number of components of a comprehensive placement system. The general thrust of this effort is to: (1) provide comprehensive occupational information to the applicant; (2) instill realistic expectations; and (3) assist the interviewer in structuring the interview with the prospective apprentice, so as to elicit information that will be useful in making the necessary employment decisions.

A major objective is the design, development, demonstration, and implementation of improved occupational counseling and job information provision capabilities. These elements of the PJM system will be programmed to operate on standalone microcomputers, the IBM PC. The system software will be designed for user friendliness, enabling use by even the most computer naive persons.

An on-line preference instrument will elicit responses from the applicant regarding such things as desired working conditions in general, as well as interest in specific trades. An occupational information database will include comprehensive information on each of the shipyard jobs. Presentation of job information will be by computer screen and/or printout.

Augmenting this microcomputer-based system, and addressing the second thrust of the effort, will be audiovisual materials that are resident on videocassette recorders (VCR). These will include a general orientation film depicting shipyard work and working conditions, and a 3-5 minute Realistic Job Preview audiovisual program for each job represented in the occupational information database.

The third major task is the development of a structured employment interview protocol. The research literature suggests that the interview process can be improved by imposing upon it a structure designed to be harmonious with the employment situation of the particular organization, in this case the shipyard.

The protocol will serve as an overall guide for the employment interview specifically and the placement process generally. It will solicit aptitude and interest test scores from extant records, as well as soliciting information directly from the interviewee. Development of the interview protocol will be based upon direct observation of current interviewing procedures, plus consultations with shipyard personnel managers who establish the policy and purposes of the interview program. It will be designed to retain control of the process for the manager, while tailoring the various procedures to the job applicant.

The Employment Scenario

Using the structured interview protocol, the interviewer will be aided in eliciting needed personal information from the applicant and in assessing applicant characteristics. The protocol will also govern the timing and scope of the delivery of occupational information to the applicant.

The job prospect will be presented with a balanced account of the realities of the shipyard work milieu via a brief Shipyard Trades Orientation Film. This orientation film will also review the benefits of employment and career possibilities. Didactic, or counseling, video screen presentations on the microcomputer will stress the importance of choosing a job on the basis of both personal and organizational considerations. The on-line preference elicitation instrument will be administered during the applicant's session on the microcomputer.

Jobs to be considered by an applicant will be selected on the basis of the structured interview and the results of the preference elicitation instrument. This will result in a focused or constrained list of jobs. Occupational information will be only be presented on those jobs that are on the applicant's focused list. For example, jobs which the applicant would not accept (based on preferences), or jobs for which there are no openings, will not appear on the list and will therefore not be considered. The RJP segments shown will be those corresponding to the applicant's constrained job list.

The net result will be a rapid focusing of applicant and interviewer attention on those available jobs for which the individual was qualified and which met the conditions important to the interviewee. This procedure will eliminate unnecessary consideration of those jobs that would be precluded by unavailability or applicant preferences. Furthermore, this procedure will narrow the range of career information that needs to be presented, making the interview less time consuming and avoiding unfulfillable employment expectations.

At the conclusion of the job preview, the applicant will have a constrained list of perhaps 3 to 5 jobs. These will then be input into an assignment algorithm along with such things as aptitude and interest scores and perhaps the interviewers classification recommendation. The algorithm referred to here is also being developed by NAVPERSRANDCEN, and will seek to optimize placement by considering job openings and the applicant's qualifications relative to those of other applicants.

NAVPERSRANDCEN will design, develop, and pilot test each component of this placement system. Certain components will be developed to the point of a prototype or demonstration version, enabling management decision on whether to proceed to full scale development. The structured interview protocol will be refined subsequent to pilot test and management review, and brought to full scale development. In addition, an initial training session will be conducted for managers who will be using the interview protocol, and the necessary training materials for future in-house training programs developed and delivered to shipyard management for use in future training programs.

Conclusion

When fully developed and implemented, this placement system should conduce to better informed employment decisions by both the organization and the individual. This system will provide adequate occupational information with which the applicant may make an informed job decision from among alternative openings, as well as materials focused on deselection by the job applicant where job preview indicates probable dissatisfaction. More realistic job expectations will be instilled in prospective workers. More complete information input into the classification algorithm will lead to more optimal placement in which both organizational and individual characteristics are considered. The end result should be enhanced placement, conducing to increased productivity and improved job satisfaction, as well as increased worker tenure and lessened turnover.

References

Baker, H. G. (May 1983) A person-job matching system for Navy recruiting: Background and needs assessment (NPRDC TN 83-7). San Diego: Navy Personnel Research and Development Center.

Baker, H. G. (July 1985) The Unwritten Contract: Job Perceptions. Personnel Journal, 64(7), pp. 36-41.a

Baker, H. G. (August 1985) Antecareer Crisis: Military Recruiting and the Youthful Job Applicant. Armed Forces and Society, 11(4), pp.565-580.b

JOB KNOWLEDGE INDICES AS A FORM OF
COST-EFFECTIVE REALISTIC JOB PREVIEW

Charles E Bethell-Fox
SP(N), Ministry of Defence, London, UK

Abstract

The Royal Navy has been investigating the effectiveness of Job Knowledge Indices (JKI's) as a form of inexpensive Realistic Job Preview in the recruitment of Navy officers. JKI's consist of true/false questions sampling aspects of Navy life such as selection, training, job content, and life-style. By their answers to these questions, enquirers reveal any areas of ignorance to Recruiting Officers who can then provide information to fill the gap. This is a cheap way of providing relevant information to candidates. The Western Recruiting Region of the United Kingdom has been involved in a trial of the JKI's and the results, based on subjective assessments by prospective recruits and Recruiting Officers, indicate that JKI's are effective. Currently, JKI's are available for Royal Navy, Fleet Air Arm, Royal Marines, and Women's Royal Navy Officers.

Introduction

Accumulated research on Realistic Job Previews (RJP's) indicates that RJP's increase "self-selection, organizational commitment, job satisfaction, performance, and job survival" (Premack & Wanous, in press). This is important especially given that without RJP's job seekers often know little about positions for which they apply. Overcoming this ignorance may explain some of the effectiveness of RJP's. However, RJP's are time consuming both for job seeker and employer, may involve extensive travel, and can be expensive especially when the number enquiring about job possibilities is large. These conditions exist in the recruitment of potential Royal Naval officer trainees in the UK and so the possibility of providing large numbers of enquirers with job relevant information quickly and cheaply has been investigated. The avenue pursued and reported in this paper is one of providing enquirers with questionnaires testing pertinent knowledge of aspects of life as a Naval officer such as selection, training, job content, and life-style. Enquirers' answers to these questions are used by Recruiting Officers to identify enquirers' areas of ignorance and Recruiting Officers then structure initial interviews with enquirers so as to provide information in these areas.

Kirton (1976), using results from the Job Knowledge Questionnaire, has shown that secondary school students are remarkably ignorant of the nature of a wide variety of jobs of the sort to which they might aspire. Our concern, using similar questionnaires called Job Knowledge Indices (JKI's), has been to see whether or not the questionnaires can function as counselling instruments to provide a type of inexpensive RJP.

Method

Subjects. Three male and two female Royal Navy officers acted as interviewers and were professionally employed as such on a two year tour of duty.

Ninety enquirers, almost all male aged between 15 and 25 years, were interviewed.

Materials. Different JKI's were constructed for Royal Navy Officer (RNO), Fleet Air Arm (FAA), Royal Marines (RM) and Womens Royal Navy Service (WRNS), containing 44, 31, 32, and 27 questions respectively. The RNO JKI covered entry requirements, selection, training, job content and life-style for RN officers generally whereas the FAA, RM, and WRNS JKI's, while focusing on similar areas, contained questions especially pertinent to those particular branches of the Service. All questions were "True"/"False" with an intermediate category for "Don't Know". A sample question similar to those actually used follows:

Naval officers:

a. spend most of their time ashore

b. are given a new appointment every two years

c. may be fired if there is surplus manpower

Administration. In most cases enquirers wrote to the recruiting office expressing an interest in becoming an officer in the Service. An interview, designed only to provide information and advice (not to select) enquirers was arranged with one of the five interviewers. In other cases enquirers simply walked off the street into the recruiting office and in one case initial contact was made during a visit to an educational establishment. In all cases JKI's were completed by enquirers prior to an interview. Interviewers marked the JKI's before the interview and during the interview, in addition to the regular interviewing schedule, discussed areas of enquirer ignorance revealed by the JKI's. Enquirers were asked to complete an anonymous questionnaire regarding efficacy of the JKI's and, if time permitted, the interviewer also completed a corresponding questionnaire.

Results

The main purpose of the study was to gauge the perceived effectiveness of JKI's from both the enquirers' and interviewers' points of view. The results clearly showed that both the enquirers and the interviewers found the JKI's very helpful. On the 90 returned enquirers' questionnaires, four questions focused on this issue. In response to the question "Did the JKI's help you discuss issues important to you?" 90% responded affirmatively and 10% negatively. Second, those responding affirmatively were asked "Which issues?": Entry requirements, selection, training, job content and life-style were mentioned as discussed issues on 78% of occasions with training, job content and life-style listed most frequently. An additional 17% simply described the JKI's as generally helpful. Third, when asked what effect JKI's had on "... your intention to apply for a commission" 30% said "Greatly encouraged", 50% "Encouraged", 17% said "No effect" (most because they had already decided to apply for a commission) and only 3% said "Discouraged". Fourth, when asked for the reasons for the effect of the JKI's on intention to apply, the vast majority of those encouraged or

greatly encouraged said that because of the discussion resulting from the JKI's they now felt they knew much more about the Service. Several also mentioned a more positive attitude towards the Service as well as increased confidence and determination about joining, suggesting that knowledge, imparted to enquirers during the interview as a result of the enquirers' responses to the JKI's, had a positive effect not just on cognitions but also on motivational processes.

Interviewers' assessments of the efficacy of JKI's were generally in agreement with enquirers' assessments. When interviewers were asked whether areas discussed from answers to the JKI's "... increased the value of the interview" 27% said "A great deal", 68% said "Somewhat" and the remaining 5% said "No effect". None said the JKI's had decreased the value of the interview. Furthermore, when interviewers were asked "Which areas were discussed from answers to the JKI's?" entry requirements, selection, training, job content and life-style were listed on 83% of occasions with 15% describing the JKI's as generally helpful.

According to both enquirers and interviewers, then, JKI's were perceived as beneficial. There was also some suggestion that this was especially true for younger enquirers. For example, those who said that the JKI's had helped them to discuss important issues were, on average, almost two years younger than those responding negatively. A similar result held for those who had been encouraged or greatly encouraged to apply for a commission as a result of using the JKI's. Particular interviewers also gave especially favourable assessments of JKI effectiveness. According to data from the interviewer questionnaires, discussion of issues raised by the JKI's typically took 10 to 15 minutes yet the interviewer who took longest, while commenting that the time consumed was worrisome, gave the second most favourable assessment of the JKI's. Indeed, over all interviewers, those taking more than 10 minutes on the JKI's were three times more likely ($p<.05$, Fisher's Test) to indicate that JKI's increased the value of the interview a great deal.

It also seems that particular JKI's can be designed to serve particular purposes. Proper choice of a specialization within the Service is important and some evidence suggested that JKI's designed to focus discussion on particular specializations did just that. For example, according to questionnaire data those enquirers who took Fleet Air Arm JKI's, as well as being more likely to have discussed Pilot versus Observer (Flight Officer) roles, were about six times more likely than other enquirers to have considered selection procedures (potential R.. aircrew Officers take specialized selection tests).

Conclusions

According to the subject assessments (ratings) reported above, JKI's are perceived as beneficial to the recruitment process by both enquirers and interviewers. This seems to be because use of JKI's elicits the provision of information to enquirers that (a) enquirers do not already know and (b) is especially relevant to forming a realistic impression of the job (c) enhances the motivation of those self-selecting to take their application further. The present exploratory research then suggests that JKI's

(which are inexpensive) can function effectively as a type of Realistic Job Preview. Future research might focus on objective rather than subjective assessments of changes in enquirers' knowledge and the effect of this knowledge on subsequent job satisfaction, performance, and survival.

References

Kirton, M (1976) Career knowledge of sixth form boys London: Careers and Occupational Information Centre.

Premack, S L, and Wanous, J P (in press). A meta-analysis of realistic job preview experiments Journal of Applied Psychology.

Acknowledgements

John Hodgkiss of SP(N) conducted some of the earlier parts of this study. The Occupational Research Centre, Hatfield Polytechnic, UK, assisted in the construction of the JKI's. Perpetua Faloon aided in the data analysis.

EFFECT OF ENTRY CHARACTERISTICS ON
SUBSEQUENT ADVANCEMENT IN A STRUCTURED CAREER**

David P. Lohmann*
United States Air Force

Abstract

The career patterns of 208 subjects entering Air Force pilot training were followed for twelve years. Initial training performance and behavioral variables were measured and correlated to subsequent attrition and performance. Significant relationships were found between self-esteem and role congruity and later attrition. Training performance, as measured by flight examination, was significantly related to advancement to senior aeronautical positions, and the speed of Air Force promotion and advancement. Career failures were correlated with peer ratings during training. Perceived future dissatisfaction with an Air Force career was significantly related to later success and career commitment. Instructor pilot predictions of student success as an Air Force pilot were significantly correlated to flying success but not career advancement. Inferences are drawn from these results about theories relating early career attributes and long term career performance.

Introduction

The military environment was used to conduct exploratory work on the theory that high self-esteem and role congruity would allow goal directed activity conducive to low attrition, high performance and satisfaction. Attrition was examined previously (Lohmann and O'Connell, 1983). The current research focuses on performance over a twelve year period. The theory of role congruity and its effect on job satisfaction, motivation, performance and attribution has been developed and applied to work situations by previous investigators. See references in Lohmann (1973).

Vocational choice theories also help explain the dynamics of role congruity and performance in a new occupational setting. A person may revise his or her choice and leave, or make a series of iterative choices to get closer to successful implementation of his or her self-concept (Amatea, 1975). If the person finds role congruity he or she will make the required commitment and acquire the required skills, stabilize in the vocation, and consolidate his or her status and seek advancement (Crites, 1969). The role of fighter pilot is suggested as the Air Force role model which new entrants need to emulate. Ashman and Telfer (1983) showed that fighter pilots can be discriminated from the general community by four EPPS scales. Previous research (Lohmann, 1973) showed that the fighter pilot role is differentiated statistically from other pilot roles. Historical work has also identified unique favorable and unfavorable attributes when facing the stress of combat. In World War II the best defense against the anxiety caused by the trauma of combat flying was a well-integrated, secure personality with a healthy ego as defense against potential threats and

*The author is on active duty in the United States Air Force. The views expressed in this paper are his own and do not necessarily reflect those of the Department of Defense or the United States Air Force.
**Presented at the National Conference Association of Human Resources Management and Organization Behavior, November 17-20, 1985.

dangers. When describing fighter pilots, more recent studies of successful NATO pilots have found characteristics similar to those of successful World War II combat pilots. The characteristics valued by the Israeli Air Force in pilot screening are similar. The fighter pilot role, with its set of historically validated desirable characteristics, has been the "sent" role model in the Air Force since its inception in 1947 (Johansen, 1979).

The hypotheses tested is that success in the subject's Air Force flying roles is directly related to how closely the individual's self concepts match that of the idealized Air Force fighter pilot. Moderating variables of aptitude, initial demonstrated skill, and satisfaction with career choice are also examined.

Method

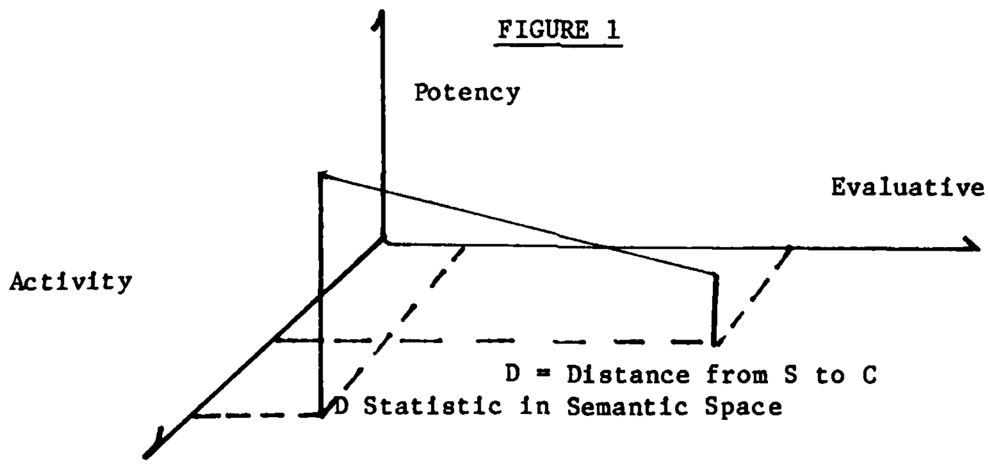
Subjects

The study was carried out among 208 Air Force officers undergoing Undergraduate Pilot Training (UPT) in 1972. Sixty-six of the subjects remain on active duty. Twenty-one other subjects failed to complete UPT. The remainder have resigned from the Air Force after serving as pilots for between four and eleven years. Previous research compared characteristics of those resigning and those continuing in the Air Force (Lohmann and O'Connell, 1983).

1973 Measures

Self esteem and role congruity were measured using the semantic differential technique validated and reported in previous research (Lohmann, 1973). It was selected to try to overcome the measurement difficulties encountered when using undimensional role concepts (most recently reported in Tuma & Grimes, 1981). Euclidian space distance (D) is measured from the origin (O) to the loci of the concept myself (S) to quantify self esteem and from the loci of the occupational personna (C) to the loci of the concept myself (S) to measure role congruity. The occupational personna selected is either fighter pilot or line pilot, the occupational personna equivalent to the type of pilot the student would eventually become, e.g., transport pilot or instructor pilot. The loci of these occupational role models in three dimensional semantic space was determined through samples of operational pilots within that particular specialty. Figure 1 illustrates this measurement concept.

FIGURE 1



Protocol analysis of a team of experienced pilots determined the quantitative translation of performance. Flying performance during training was measured by assignment of numerical grades (4.0 - 0.0) to check rides. The peer rating was the normalized mean value of the rank ordered merit assigned by fellow flight members. Flight size varied from 16 to 28. Pilot aptitude was taken from the Air Force Officer Qualifying Test. The future perceived dissatisfaction index was measured using the Porter and Lawler satisfaction survey instrument. The commander's assessment of an individual's future performance was the normalized rank order of each flight commander's students.

1985 Measures

Flying performance was measured by awarding a unit of merit for the assignment to certain flying positions which required selection in competition against peers and resulting award of a high skill identifier.

Achievement in professional military education was measured by completion of the junior, middle management and senior leadership schools. Deferment was equivalent to the number of times an individual was evaluated by a promotion board and not selected. Poor effective ratings was the number of times the subjects were assigned less than a perfect score on the annual effectiveness report. Years to make major and years to regular commission are time elapsed from date of commission.

Results

Table 1
Cross Correlations of 1972 Training Variables and 1985 Performance Variables

	1985 Performance Variables						
	FLY	TIC*	PME	GR	PASS	MAJ	REG
M.	1.92	53.3	2.87	3.83	.39	13.32	6.27
SD.	1.04	23.6	.89	.48	.91	3.25	5.23
1972 Training Variables							
Self-Esteem	.05	.06	.24*	.22	-.12	-.20	.08
Fighter Pilot Role Congruity	-.09	.11	-.12	-.23*	-.04	-.30	-.16
Actual Pilot Role Congruity	-.07	-.19	-.04	-.24*	-.01	.01	.02
Check Flight Performance	.27*	.29*	-.03	-.23*	-.06	-.20*	-.27
Class Standing	.04	.24*	-.16	.07	.08	.04	.18
Peer Rating	.21	.19	.11	.27*	-.21*	-.35*	-.16
Commander's Assessment	.24*	.12	.10	.12	.20	-.15	.06
Perceived Future Dissatisfaction	.07	.17	-.31*	-.22*	.14	.13	.22

*p $<$ .05
*Time in Cockpit In Percent

Table 1 presents the cross correlation between 1972 training variables and 1985 performance variables together with 1985 means and standard deviations. The self-esteem measure shows correlation to professional military education achievement and current grade. The correlation with professional military education may be attributable to the member's confidence in his ability to succeed in these schools and therefore a propensity to attend early or take the course by correspondence while accomplishing his normal duties. The significant correlations of self-esteem and role congruity to grade but not to the number of years to be promoted to major indicates that these variables may be more a result of maturity than a success predictor. Check flight performance has significant correlations with all measures of future performance. These check flights not only measure a student's ability to perform aeronautical maneuvers but also to operate in accordance with strict procedures and to function under stress. These abilities to do well on these requirements may explain subsequent success as a military officer. Similar results were found in the Royal Air Force (Elshaw & Ledderdale, 1982). Class standing is only related to one variable, the percentage of one's career spent in the cockpit versus staff officer or management positions. Peer ratings have the strongest set of correlations to subsequent performance and failures. In the self comparisons inherent in peer ratings, students may be accomplishing a more comprehensive evaluation of peer attributes than any single psychological measure can match. In the competitive closed promotion system of the miliary, this comparative assessment may be more valuable than in less structured settings. Subjects with the lowest perceived future dissatisfaction with an Air Force career are those willing to make the institutional commitment of attending a military school. Also it appears that the subjects with time in the Air Force before attending UPT had more information about the career and had lower perceived dissatisfaction. A similar effect can be seen in the relationship between this variable and the time to a regular commission.

Discussion

Role congruity measurement difficulties outlined by Levinson (1971) appear to continue. It is apparent from the relationships to twelve year future performance of three subjective evaluations, flight examination scores, commander estimates of future operational performance and peer ratings, that there is some global characteristic of competence and potential not yet quantifiable. Previous research (Lohmann, 1973) (Lohmann and O'Connell, 1976, 1983) also indicated the undimensional characteristics of aptitude, intelligence, effort, manifestation of apprehension and academic performance were poor indicators of this enduring global characteristic.

As suggested by Tziner (1983) in research on Israeli Air Force subjects, early personality measures may have good explanatory power when assessing commitment to a long term career occupation but, given the adaptability of the individual, may have only limited utility in explaining long term career achievement. This conclusion is also supported by attempts to develop predictors. Regression analyses were performed with all 1985 dependent performance variables and all 1972 training measures. In no case were significant improvements achieved over bivariate correlations. It appears that career patterns are complicated chains, each

step dependent on previous paths. While self-esteem and role congruity measures may be useful short-term concepts for predicting early commitment, prediction of long performance based upon these behavioral measures does not appear fruitful. The measures account for only a small portion of the differences in individual performance. More lucrative predictors appear to be task competence, expert evaluation, and the subject's adaptive ability to move through the career maze by mastering each step; step by step. These results lend support to the theorists who propose that vocational development is a series of iterative choices.

References

Amatea, Ellen S. (1975) Contribution of career development theories. (Readon, Robert C., and Burck, Harman P. (Eds.) Facilitating Career Development. Springfield, IL: Charles C. Thomas.

Ashman, A., and Telfer, R. (1983) Personality profiles of pilots. Aviation, Space and Environmental Medicine, 54(10), 940-943.

Cassie, A. et al. (1964) Aviation Psychology. Paris: Norton & Co.

Crites, John O. (1969) Vocational Psychology. New York: McGraw Hill.

Elshaw, C. C. and Lidderdale, I. G. (1982) Flying Selection in the Royal Air Force. Newsletter of the International Test Commission, 17, 3-13.

Johansen, William A. (1979) Characteristics and skills of successful fighter pilots: A better method of selecting Air Force pilots, Research Report, Air Command and Staff College, Maxwell Air Force Base, AL.

Levinson, Daniel J. (1971) Role, personality, and social structure in the organizational setting. (In Richard R. MacDonald and James A Schellenberg (Eds.) Selected Readings and Projects in Social Psychology. New York: Random house, Inc.

Lohmann, David P. (1973) An examination of some behavioral correlates of Air Force Undergraduates Pilot Training through the use of the Porter and Lawler Performance/Satisfaction Model. DBA Dissertation, Arizona State University.

Lohmann, D. P. and O'Connell, M. J. (1983) The relationship of role congruity to Air Force pilot turnover and performance. Southwest Academy of Management Journal.

Tuma, N. B. and Grimes, A. J. (1981) A comparison of models of role orientations of professionals in a research oriented university. Administrative Science Quarterly, 26, 187-206.

Tziner, Aharon (1983) Choice and commitment to a military career. Social Behavior and Personality, 11(1), 119-128.

Early Career Experiences of Air Force Academy Graduates

Sharon C. Conley
Organizational Analysis and Practice, Inc.
and
Frank R. Wood
U.S. Air Force Academy

Abstract

This paper examines written comments of young officers to identify factors which enhance or debilitate careers. Implications for Academy and Air Force policy are discussed.

INTRODUCTION

It is well established in the career development literature that early occupational experiences of individuals are highly predictive of later job satisfaction, job adjustment, and overall organizational commitment (Hall,1976). It was with this observation that we began a study of the career experiences of recent AF Academy graduates. The purpose of the study was to examine various aspects of the way recent Academy graduates perceive their Air Force jobs.

This study was designed to focus on three sets of job-related variables -- job expectations, job characteristics and critical psychological states-- all related to job satisfaction (Hackman and Oldham, 1975). The quantitative data, collected by a mailed questionaire, related job characteristics to job satisfaction, motivation, and psychological involvement in one's work (Conley,1983). In addition to the quantitative data, qualitative data in the form of written comments identified early career experience which enhanced or debilitated young careers.

The purpose of this paper is to present our findings regarding these early career experiences. First, we will present eleven observations formulated from the qualitative data provided by the respondents. Where appropriate, the career enhancing or career debilitating nature of these observations will be discussed. Finally, we will discuss the implications these observations have for current operation of the Academy or the Air Force.

METHOD

The subjects included all the female graduates of the Air Force Academy Class of 1980 and a matched sample of males in the same class. Of the 172 lieutenants who were mailed surveys, 110 (63% including 53 males and 57 females) returned useable questionaires. Nearly 40% of the sample were in flying jobs (aircrew members) and the rest were in support specialties.

FINDINGS

The following eleven observations are based on the content of written comments offered by the respondents and the findings

from the earlier quantitative analyses of data (see Conley, 1983).

1. Before they graduate, cadets expect that their future Air Force jobs will provide them with the opportunities to: use a variety of skills, receive job feedback, complete tasks from beginning to end, and work on tasks that are significant. However, they expect only moderate levels of job autonomy.

On average, cadets' perceptions of their future Air Force jobs do not differ significantly from those of professional-technical workers in general (Conley, 1983). The only exception to this was with respect to autonomy; cadets had expectations of job autonomy which were lower than the average self-reported autonomy of professional-technical workers. This finding provides some reference for the understanding of reported levels of autonomy.

2. Academy graduates enter jobs that meet their expectations, with one exception: support officers are in jobs that provide them with more autonomy than they expected.

The quantitative analysis found no significant differences between cadets' job expectations before graduation and actual job characteristics found in their officer jobs execpt with regard to autonomy. While all officers in this sample reported more autonomy than they expected (4.7 on a 7 point scale vs 4.1 expected) the discrepency was greatest among support officers who reported much higher actual autonomy in their jobs (5.1 vs 4.2 for flyers). This finding raises some important issues: what is it about Academy training that makes some of the cadets' expectations very realistic and others not (e.g., autonomy); and why was the reported autonomy for flyers and support offices different? The first issue addresses the effect of the cadets' own limited autonomy on their perceptions of Air Force jobs and the relatively successful effort of the faculty to teach the cadets what to expect when they graduate. The second issue supports other research which suggests support officers do have more autonomy in their jobs. (Wood, 1981)

In general, we expect that greater levels of autonomy than expected would be career enhancing experience. The actual effect will be examined in another study.

3. While support officers report more autonomy than do flying officers, flying officers report more skill variety and feedback -- and experience their jobs as more meaningful -- than do support officers.

Flying officers in this sample were significantly more likely than support officers to report their jobs were "very meaningful" (Conley, 1983) probably because flying is seen as the core activity of the Air Force. The the flying officer's perception of greater skill variety and feedback can be attribut-

ed to the fact that much of their early flying career involves learning new skills and their job performance is reported to them clearly and precisely by the flight instruments of the aircraft.

4. Early career Academy graduates are highly motivated to perform their jobs.

The mean score for "internal motivation" reported by these officers was 5.9 (on a 7 point scale). This observation is important because it suggests the best personnel strategy for this group is to continue providing career enhancing jobs, to focus on sustaining high levels of motivation, rather than fixing the lack of motivation often assumed of new personnel.

5. Lack of/presence of job characteristics - autonomy, skill variety, task identity, feedback, task significance can be a debilitating/enhancing force in early careers.

Jobs which provide challenge, autonomy, skill variety, task identity, feedback and task significance, lead to positive career experiences. As one cadet commented, the reverse is also true:

"I think the major problem...in adjusting to the Air Force is that I am not challenged enough. At the Academy I was busy all the time and constantly pushed. Now I feel like I want to work harder and learn more and I am not really given the opportunity."

6. Unsatisfactory working relationships with others can be a debilitating force in early careers.

Support and guidance from commanders/supervisors can be career enhancing when young officers say ...
"my (supervisor ratings) are better than I'd ever expected based on my performance at USAFA (and) I...(now) have got confidence in myself..." or debilitating when ...
"(commanders) do not expect second lieutenants to perform."

Women graduates indicated that problems with male co-workers can be debilitating, although their comments suggest that they have learned to accept and cope with these problems constructively.

7. Disappointment/satisfaction concerning career fields or assignments can be a debilitating/enhancing force in early career.

First assignments are critical. Comments such as:
"I feel like perhaps I would find greater job satisfaction in a different career field... I have not been given a chance so far to exercise the education I received... I don't feel like I am being utilized very effectively..." suggest dissatisfaction and despair.

8. Graduates tend to differentiate between satisfaction with a job assignment and satisfaction with the Air Force as a career.

Some of these lieutenants report that while they do not like their present assignment, they see it as contributing, in the long run, to a satisfying Air Force career. Others point out that while they like their current job they do not see it as contributing to their Air Force career. Typical comments were: "I hate my present job... but...it will be a very good stepping stone to any other career field that I want. So... I will like it in my records in the future" or "... being a first assignment IP is a nice job but... it limits future assignments, many people in my situation are initially bitter."

9. The Academy faces a major delimma in the descrepant requirements of giving cadets a sense of the general Air Force mission and the specialized knowledge they will need in their future jobs.

Some cadets react negatively to what they perceive as an undo emphasis on depth or breath in their training: "The academy prepares you for being a general officer, not near enough for being a junior one..." or "USAFA is set up solely for the pilot (graduate). It was an incredible let down to discover how poorly ...myself...(and) my classmates were prepared for non-flying assignments."

If cadets are not prepared for the jobs they actually encounter their early experience can be debilitating.

10. The Air Force faces a delimma posed by the requirement to give early career experiences that provide substantial autonomy and the need to maintain some control over the work activities of young officers.

There is a negative impact when work activities are seen as too tightly controlled or performance standards too flexible. For example: "No one expects you to do well as a second lieutenant. They have set some of the lowest possible standards of performance as a result... commanders...expect less..." and "Two of the hardest things to deal with is the lack of feedback and the lack of structure."

11. Early career decisions of graduates are tentative at best.

The overwhelming majority, when asked about their career intentions, responded that they would "wait and see." A current follow-up study of these cadets at the "five-year" mark will explore some of the factors affecting their decisions.

IMPLICATIONS

The qualitative comments of these graduates suggested some issues that were not considered when the quantitative study was designed. The issues raised are significant in that they affect the career decisions of these expensive manpower resources and they carry implications for changes in the operation of both the Academy and the Air Force.

If the Academy can't -- and isn't set up to -- give cadets a "preview" of specific, operational aspects of their future jobs, we need to be concerned with finding different ways of smoothing the transition from the Academy to Air Force jobs, e.g., orientations, supervisory training, long-range career planning.

Job status was not directly addressed in this study, but the findings concerning differences between perceptions of flying and support jobs raise some important questions about how job status is conceptualized. It was found that officers in flying jobs perceive their work as more meaningful, but officers in support jobs have more autonomy. Job "status" -- from the orientation of the flyers -- appears to reside in the closer fit between the job and the Air Force mission; "status" from the perspective of support officers lies in their greater freedom and on-the-job discretion. Instead of thinking about jobs as having their status defined by either autonomy or meaningfulness, we need to begin to find ways of enhancing the autonomy and meaningfulness in each job, at each career stage thereby enhancing the career growth and commitment of these key Air Force officers.

REFERENCES

Conley, S.C., The Relationships Among Expectations, Perceived Job Characteristics and Work Outcomes. Unpublished doctoral dissertation, University of Michigan, 1983.

Hall, D.T. Careers in Organizations. Pacific Palisades, Ca.: Goodyear, 1976.

Hackman, J.R., and Oldham, G.R., Development of the Job Diagnostic Survey. Journal of Applied Psychology, 1975, 60, 159-170.

Wood, F.R. Air Force Junior Officers: CHanging Prestige and Civilianization. Armed Forces and Society, 1980, 6, 483-506.

Roles of the Mentor
as Perceived by the Mentor and the Protege

Benjamin L. Dilla, Captain, USAF
Air Force Institute of Technology

Interest in the topic of mentoring has blossomed in recent years although empirical research is virtually nonexistent. Many authors have discussed functions or roles of mentor in developing their proteges, but no one has empirically examined the actual use or relative importance of these roles. This research tested the comprehensive list of ten roles identified by Lea and Leibowitz (1983). Over a two year period, 347 officers attending Air Command and Staff College and Air War College were surveyed for their views on the mentoring process. In this paper, results from the protege's perspective are contrasted with those from the mentor's perspective. Although all ten roles were identified as being played by at least some officers, mentors tended to see themselves in more active and directive roles than those reported by proteges. Rationale for these findings and suggestions for future research are offered.

Mentoring can be defined as a relationship between a senior member and a junior member of an organization in which the senior member is influential in molding and shaping the career of the younger member (Uecker & Dilla, 1984). The concept of mentoring has recently received considerable attention throughout the field of management. Trade journals, in particular, abound with articles ranging from cross-gender mentoring to reasons why one should (or should not) enter into a mentoring relationship.

Recent interest in mentoring was spurred by Jennings' (1971) book Routes to the Executive Suite. In his discussion of executive and managerial development, Jennings emphasized the importance of a "sponsor" or individual who gave an inordinate amount of counseling and guidance. He claimed that most company presidents have had a sponsor. He felt this was particularly true in organizations that promote from within since one would need a sponsor to become a member of the "ingroup" (Jennings, 1971, p. 145).

Following this book, a wealth of articles began appearing in the popular literature through the late 1970s and into the 1980s. Books and articles approach the topic from many different perspectives, but most present only opinions based on personal experience or case studies of organizations which have employed mentoring.

A serious problem is that mention of the word mentoring in many circles, including the military, typically elicits visions of sponsorship, careerism, and the easy way to the top. However, researchers have approached it simply as one form of interpersonal relationship that may influence a person's growth during a career (Kram, 1985; Lea & Leibowitz, 1981). From this perspective, the mentor does not provide a free ride to the top but presents a positive example, counsel on the artful skills of leadership, and opportunities in line with the person's capabilities and potential. Without empirical evidence, the difference in perspectives remains strictly a matter of personal opinion.

Empirical work in the area is limited to a large-scale survey (Roche, 1979) and more recent interviewing studies by Kram (1985) and her associates. Roche's (1979) study was a survey of 1250 business executives. His results supported the prevalence and positive effects of mentoring in developing leaders in American business. This study found that "nearly two-thirds of the

respondents reported having had a mentor or sponsor" (Roche, 1979, p. 14). The research also discovered that mentored executives earned more money at a younger age and had a higher degree of satisfaction with their jobs and their career progress. Furthermore, those in the mentored group were better educated and more likely to have formulated and followed a career plan.

More recently, Kram (1985) increased the understanding of the mentoring process through in-depth interviews of mentor-protege pairs. Her research centered around the satisfaction of individual needs at various career stages but also showed positive effects of mentoring on the leadership development and career progression of proteges.

Even the most recent literature is still struggling to define exactly what a mentor is and does. From her research and some earlier works, Kram (1985) derived a list of nine functions which might be observed in a mentoring relationship. Her list included five career functions (sponsorship, exposure-and-visibility, coaching, protection, and challenging assignments) as well as four psychosocial functions (role modeling, acceptance-and-confirmation, counseling, and friendship). Earlier, in one of the most comprehensive definitions of the mentoring concept, Lea and Leibowitz (1983) discussed ten roles or behaviors of the mentor. These behaviors include teaching, guiding, advising, counseling, sponsoring, role-modeling, validating, motivating, protecting, and communicating. Their list appears to encompass all the ideas given by others in the field.

Goals of the present research were to apply Roche's empirical (survey) approach to examine mentoring in the U.S. Air Force and to expand the survey to examine the relative importance of various proposed roles of the mentor.

A modified and expanded version of Roche's survey was applied in the Air Force in two recent studies; both were completed under my direction as master's thesis projects at the Air Force Institute of Technology. Captain Michael Uecker first surveyed 252 officers attending Air Command and Staff College (ACSC) and Air War College (AWC) to examine the prevalence and effects of mentoring in the Air Force. Based on these findings, Captain Francis Lewandowski surveyed 95 AWC officers to gain insight into their perspective as those who had been mentors as well as proteges. Some results of these studies have been presented elsewhere (Lewandowski and Dilla, 1985; Uecker and Dilla, 1984); the present report highlights the data concerning roles of the mentor and focuses on the contrasting perspectives of mentors and proteges provided by the two studies.

Methodology

Study 1

Sample. Surveys were sent to 160 active duty Air Force officers attending AWC and 210 at ACSC in the spring of 1984. This sample was chosen because these officers had demonstrated high potential and were slated for future positions of leadership. Of the 370 officers who were sent surveys, 252 (68.1%) responded.

Procedure. The surveys were mailed to a central point at Air University headquarters and distributed through the two schools to the students. Each school then collected the questionnaires and sent them back through the central office. The total time the students had the surveys was approximately two weeks. Because of the frequent negative connotation of mentoring in the Air Force, the survey cover letter defined mentoring and carried an endorsement by the Dean of the School of Systems and Logistics (a USAF O-6).

Measures. Mr. Gerald Roche provided the researchers with a copy of his questionnaire (Roche, 1979); this provided the basis for the survey instrument used. Modifications were performed to tailor the instrument for a military sample. Some items were deleted as they were irrelevant to the Air Force (for example, questions concerning salary, executive perquisites and previous employers). Sections were added addressing the roles of the mentor and concerning major command identification, military rank and promotions.

With respect to roles of the mentor, respondents were presented with a list of the ten roles discussed by Lea and Leibowitz (1983), some with one-line definitions for clarification. They were asked to rate each role as to whether it had been a "most important", "major", or "secondary" role which their mentor played, or a role "not played" by their mentor.

Study 2

Sample. Officers designated for the 1985 entering class of AWC were surveyed. These officers were lieutenant colonels and colonels (0-5 and 0-6) coming out of a variety of leadership and staff positions including squadron commanders, directors at air division level, and system program directors. Of the 112 officers sent surveys, a total of 95 (85%) responded.

Procedure. Surveys were mailed to officers at their duty addresses several months before their departure for AWC. Participation was voluntary, and respondents were assured of anonymity. They were asked to mark their responses directly on the survey instrument and return it in a postage-paid return envelope.

Measures. The survey instrument paralleled Uecker's (1984) survey, but the format was expanded so that the officers would respond to some items, including roles of the mentor, separately from the perspective of being a protege and from being a mentor.

Results

In Study 1, mentoring was not found to be as prevalent in the military sample (42% versus Roche's 64%), but the effects were similar. Mentored officers were better educated and more likely to have formulated a career plan. They were more likely to have been promoted early (a parallel to Roche's finding of earning more money at an earlier age) and had greater job and career progress satisfaction. These respondents saw their mentors primarily in the sense of role models, motivators, and advisors rather than in more directive roles such as sponsor and protector (Uecker & Dilla, 1984).

Of the 95 respondents in Study 2, 58 reported having had a mentor; this 61% rate of mentoring is not significantly different from the 64% reported by Roche (1979) for the private sector. In the second section of the survey, 46 officers (48% of the sample) stated that they had served as a mentor to another individual. Two-thirds of these officers (30 of the 46) reported that they currently had one or two proteges (Lewandowski & Dilla, 1985).

Study 2 failed to find any significant differences between mentored and unmentored groups with regard to formulation of a career plan, job satisfaction, or early promotions as had been found in the previous research. However, the small sample size of this study may have been a limiting factor. A significant difference was found between groups with respect to career

progress *satisfaction*. For the 46 officers who had served as mentors, there were no significant differences with respect to early promotions or career progress satisfaction; however, they had significantly higher job satisfaction than their contemporaries who had not been mentors.

Study 2 examined roles of the mentor from the perspective of both proteges and mentors, allowing for a contrast between the two perspectives as well as a comparison with the results of Study 1 for proteges from a similar sample. Mean responses for each role for Study 1's group of 106 proteges and Study 2's group of 58 proteges and 46 mentors are presented in Table 1. Rankings of the ten roles within each group are also presented for ease of comparison.

Table 1
Roles of the Mentor

	Average Rating[1] (& Rank within Group)		
	Study 1	Study 2	
Role	AWC & ACSC Proteges (n=106)	AWC Proteges (n=58)	AWC Mentors (n=46)
Role Model	1.924 (1)	1.840 (1)	1.711 (4)
Advisor	1.853 (2)	1.706 (5)	2.158 (1)
Motivator	1.800 (3)	1.734 (2)	1.798 (3)
Supporter	1.613 (4)	1.511 (6)	1.500 (8)
Counselor	1.598 (5)	1.708 (4)	1.932 (2)
Communicator	1.505 (6)	1.093 (9)	1.561 (7)
Guide	1.500 (7)	1.321 (8)	1.675 (5)
Teacher	1.441 (8)	1.344 (7)	1.611 (6)
Sponsor	1.426 (9)	1.716 (3)	1.343 (9)
Protector	0.964 (10)	0.713 (10)	1.095 (10)

[1] Ratings and assigned scale values were:
Most Important = 3; Primary = 2; Secondary = 1 ;
Not Played = 0.

The largest difference in ranks occurred for the controversial role of "sponsor". This term, which often carries a negative connotation in the Air Force, was rated relatively low by the sample in Study 1 and by the smaller group of mentors in Study 2. Yet, when rated from the perspective of being a protege, the respondents of Study 2 gave it the third highest mean rating; in fact, it received the largest percentage of "most important" responses for this sample.

Other notable differences included the advisor role, which received the highest mean rating for the group of mentors (also the highest absolute rating in any of the groups), but emerged fifth when rated from the protege's perspective in Study 2 and second in Study 1. In the opposite direction were the results for "role model", rated highest by both groups of proteges but only fourth among the roles rated *by the mentors*.

Discussion

Both studies helped establish that mentoring is a fact of life in the Air Force as it is in the private sector. Comments at the ends of the surveys indicated that many officers understood and supported the use of mentoring in the Air Force; however, many harsh, negative comments revealed stereotypes and misconceptions even among senior officers. We who have been involved with this research believe the Air Force should publicize the reasons for, and potential benefits of, the informal mentoring process. Furthermore, research on mentoring should continue and be expanded to both a broader scope and higher level of officers in order to better understand the dynamics and effects of the process.

Regarding the roles of the mentor, proteges most often described their mentor as a role model, although the mentors assigned less importance to this function. This difference seems natural since it would be difficult for the mentor to tell to what extent the protege is observing and striving to emulate his behavior. The most important roles from the mentors' perspective were those of advisor, counselor, and motivator, roles that are more active but still not highly directive. Proteges were in agreement with the relative importance of the motivator role but seemed less willing to admit that their mentors had to "counsel" them. This difference may be due to some of the negative connotations to the term itself.

There was clear agreement across all groups that the mentor does not serve as a protector to the protege, or at least does so very infrequently. It is noteworthy that the mentors assigned greater importance to this role than either group of proteges; further study of the occurrence of this function may be merited. Respondents also tended to play down the importance of the sponsor role, with the exception of the AWC proteges. It appears that the more senior group of AWC officers perceived that their mentors had provided growth opportunities for them to a greater extent than did Study 1's combined sample of AWC and ACSC students. However, the low rating by the same group of officers when viewing the roles as mentors is puzzling.

Further research with more detailed definitions of the terms and larger sample sizes may be the only way to resolve or clarify these differences. Again, we support and encourage further research and a more open dialogue on the topic within the Air Force.

References

Jennings, E.E. (1971). Routes to the executive suite. New York: McGraw-Hill.

Kram, K.E. (1985). Mentoring at work: Developmental relationships in organizational life. Glenview, IL: Scott, Foresman.

Lea, D., & Leibowitz, Z.B. (1983). A mentor: Would you know one if you saw one? Supervisory Management, 28, 32-35.

Lewandowski, F., & Dilla, B.L. (1985). Mentoring in the United States Air Force: The mentor's perspective. Proceedings of the 27th Annual Conference of the Military Testing Association.

Roche, G.R. (1979). Much ado about mentors. Harvard Business Review, 57(1), 14-28.

Uecker, M.E., & Dilla, B.L. (1984). Mentoring as a leadership development tool in the United States Air Force. Proceedings of the 26th Annual Conference of the Military Testing Association, pp. 423-428.

PANEL SESSION

TOTAL QUALITY CONTROL: IMPLEMENTATION ISSUES IN A NAVY INDUSTRIAL COMPLEX

Samuel B. Landau, Marilyn Monda, E. Chandler Shumate, and Joyce Shettel-Neuber
Navy Personnel Research and Development Center

Wayne Putnam
Naval Air Logistics Command

ABSTRACT

The purpose of this panel session is to describe the efforts and results of implementing a Total Quality Control (TQC) system, in a division, at a Naval Air Rework Facility. The significance of TQC methods is increasing as it is being implemented by many public and private sector organizations. A TQC system is a managerial philosophy and approach which emphasizes the production of high-quality, as opposed to high quantity, goods and services by following several principles. Briefly, TQC is process rather than product-oriented; it focuses on consumer needs; and attempts to improve quality and productivity by reducing product/service variability through various statistical techniques. This panel will discuss various theoretical and applied issues of the philosophical and structural changes associated with implementing and measuring TQC procedures.

Organizational change variables, such as awareness of the need for change, commitment to the change, communication and acceptance of the change will be discussed in light of past research and as they apply to the present industrial setting. Implementation of changes in management philosophy and behavior will be discussed as they relate to various job characteristics (such as autonomy, feedback, and satisfaction), organizational commitment, training needs, performance measures, and organizational reporting and communication requirements (vertical and horizontal). The implications and applications of this approach to the entire organization and to a larger group of organizations will also be discussed.

Panel Format for

Total Quality Control: Implementation Issues
in a Navy Industrial Complex

The session will be organized as follows:

1. Total Quality Control as an Organizational Change: Implementation Requirements
Samuel B. Landau

2. Total Quality Control: An Overview at a Naval Air Rework Facility
E. Chandler Shumate

3. Training Needs: Tools for Implementation
Marilyn Monda

4. Measurement and Organizational Impact: Effects of Implementation
Joyce Shettel-Neuber

5. Implications and Expansion of Total Quality Control in Public- Sector Organizations
Wayne Putnam

Computer-Based Job Performance Aids

Anita S. West
Denver Research Institute

Abstract

Under several current and recent contracts with the Department of Defense, NASA and the Department of Education, DRI researchers have been investigating the use of computer-based training and job performance aids, with an emphasis on the needs of high turnover occupations and those requiring frequent modifications of technical information and management procedures. This panel considers issues involved in the planning, development and implementation of these aids. Contrast is made between population constraints and presentation requirements of instruction vs. information with consideration of the strengths and weaknesses of traditional instructional design theory for both of these applications. The panel demonstrates how techniques of artificial intelligence have contributed to and changed the concepts and uses of information systems. Also addressed will be the effects of a system designer's assumptions about learning and the adult learner in the creation of both instructional and informational systems. All of these principles are explored in a questions and answer format from the perspectives of two contrasting job performance environments: the project management environment and an equipment maintenance environment.

The panel participants, whose statements follow, are conducting their individual programs at the Denver Research Institute.

Participants
Barbara L. McCombs

Cognitive Processes and Adult Learner Models

Current theories of learning view the learner as an active information processor. Furthermore, in order for learning to occur, learners must transform new information to fit their existing knowledge structures. Computer-based training and information aids must be designed to accommodate individual differences in the types and amounts of prior knowledge and knowledge structures in order to provide a meaningful context for learning and job performance. This is particularly critical for adult learners who have well established knowledge structures and prior experiences.

James P. Kottenstette

A Philosophy About Job Performance Aids

The notion that people can use computer-based knowledge resources on the job, and that such use provides a practical alternative to traditional training initiatives, is a powerful idea that promises to liberate both the individual and his or her organization: The individual is free to seek out explanation or reference materials when needed to improve job performance, and the organization is finally able to communicate consistently about its role and its expectations through the instructional and informational resources provided to its personnel.

Kenneth Harmon

Expert Systems and the Artificial Intelligence Perspective

Artificial intelligence (AI) technology will have a significant impact on the development and use of job performance aids (JPAs). The AI concept of "device models" allows for the integration of design and support activities for engineered devices. This use can ensure that support issues, such as testabililty, repairability, and technical documentation are considered during design. The device model can be used to develop the fully proceduralized JPAs useful in troubleshooting. Also, using device models, hybrid JPAs which adapt to the level of expertise of the user can be employed and embedded training become feasible.

Linda Brannan

User Interface

It is essential in any job performance aid authoring or delivery system for the designer to keep the user's needs and level of experience firmly in mind. Design objectives can provide flexible user control over sequencing of material and provide at least two levels of user/program interaction, novice and expert. The well designed system provides visual cues and reinforcements for the first- time and experienced user and helps him or her to distinguish between information for one-time use and instructional building blocks for generalizability and future transfer.

COGNITIVE STYLE AND DECISION MAKING

Elliot E. Entin
ALPHATECH, Inc.

John A. Forester
Armstrong Aerospace Medical Research Laboratory

William E. Gutowski
ALPHATECH, Inc.

ABSTRACT

In the first of two simulation experiments, subjects identified as having an analytic or global cognitive style were required to discriminate between a missile attack or a missile test condition, based on imperfect information. Those possessing an analytic style made significantly better discriminations than subjects exhibiting a global style. An additional analysis investigating whether recognition memory could be a factor in the group differences was also performed, and results showed recognition performance of analytics to be superior. The second experiment assessed two additional cognitive styles and manipulated reports of missile heat signature and missile sites attack probabilities. Results from the signal detection analysis replicated the earlier style effects. Four of the experimental factors, attack/test outcome, site attack probability, missile heat signature, and launch update reports had significant main and interactive effects on attack probability assessment.

INTRODUCTION

In a typical command, control, and communication environment the decision maker must take in information, often from various sources, deal with uncertainty associated with that information, and make a decision based on the information in a timely manner. For the two paper and pencil simulation experiments reported here, the subjects taking the role of a missile warning officer were required to discriminate between real and false missile launches and whether the missile was on an attack or test trajectory. In making this discrimination, the subjects must integrate imperfect heat sensor and X-Y position data presented on an overhead projection map of the area, determine if a missile is involved, and if so, whether it appears to be on an attack or test trajectory. The initial information is updated two additional times for a total of three reports for each event (or trial).

To assess how well subjects can make this discrimination, each experiment was studied as a signal detection problem (Green and Swets, 1966). In this case, attacks are the signal and tests or false alarms are noise. For any group of subjects, the discrimination ability can be examined as an ROC (Receiver Operating Characteristic). The ROC analysis allows a determination of whether subjects with different cognitive styles differ in their discrimination ability. Each design also provides a means of investigating multiple cue decision making in uncertain conditions: in the second study various sources of uncertainty are systematically manipulated to observe their effects on three cascading decisions.

EXPERIMENT 1

Method

A major goal of Experiment 1 was to observe if subjects identified as possessing an analytic or global cognitive style would differ in their ability to make discriminations.

Subjects - Twenty-five student volunteers from a local technical college served as subjects.

Procedures - All subjects completed the Group Embedded Figures Test (Witkin et al., 1971) as a measure of analytic versus global style (Benbasat and Dexter, 1979). Subjects were then read induction material designed to arouse their interest, induce a belief in the importance of the task, and manipulate the various factors of uncertainty in the experimental design. They were told their duties were to monitor heat sensor data and position data of the heat source, telemetered from Us's satellites to determine if the country of Them has launched an attack.

Subjects were told that a missile's heat signature ranged from 3 to 9 on a scale of 1 to 12. Uncertainty was introduced by explaining that the heat sensor data have an error of ±2 and that other sources like explosions, large fires, electrical storms can all return readings very similar to missiles. Moreover, position data also could exhibit an error radius of 10 miles. Given these elements of uncertainty, it was the subject's job to decide on which of two trajectories, test or attack, a missile was traversing. Each of Them's 10 missile sites was associated with two trajectories. One led into the country of Us (attack trajectory) and one came close to but missed the country of Us (test trajectory). Subjects viewed a map showing the test and attack trajectories for 90 seconds. All judgments of test or attack were made on maps similar to this one except the test and attack trajectories were not drawn in. The heat sensor reading and heat source position (indicated by a black dot on the map) were updated in each of three reports.

Subjects were instructed to integrate all information and then indicate their probability that Them had launched an attack. Probability estimates were made on a 0 to 4 scale. Fifteen seconds were allotted for each rating. Thirty trials of three reports each were conducted. A random half of the trials depicted an attack trajectory. Sensor readings were generated according to a modified random process; essentially, care was taken to see that the sensor intensity was reasonable given a missile or non-missile trial. Unknown to the subjects, all sensor positions were plotted along their designated trajectories with no error.

Results and Discussion

Comparison statistics for the two style groups show significant differences for percentage of area under the ROC curve (analytics' mean = 87.8%, globals' mean = 63.3%, $p < .01$) and d' (analytics' mean = 1.53, globals' mean = 0.49, $p < .01$) in favor of the analytic style. These findings clearly suggest that cognitive style differences can have important effects on multiple cue decision making.

The data of this experiment may also be viewed as indexing recognition memory for the attack and test trajectories from the various launch sites. Moreover, it is possible that a memory component can be accounting for the cognitive style differences, because subjects were required to memorize the attack and test trajectories. Thus an estimate of memory storage, Θ_{-s}, was developed which assumed that strength of memory traces can be treated as a dichotomous variable. That is a trace was either sufficiently or insufficiently stored. If a memory trace is sufficiently stored, then a subject will always give the maximum probability response of 4 or the minimum response of 1; if it is not sufficiently stored, then a subject's response will reflect rating biases and/or guessing.

An ANOVA of the memory storage measure, Θ_{-s}, shows significant main effects for cognitive style and reports. That is, analytics demonstrated larger Θ_{-s}, than globals ($F = 5.91$ $df = 1,17$, $p < .03$), and Θ_{-s} increased with each additional report update ($F = 16.34$, $df = 2,34$, $p < .001$). The interaction did not, however, reach significance. An *a priori* contrast computed for reports 1 and 3 (i.e., report 2 was omitted) by cognitive style did, however, prove significant ($t = 1.80$, $df = 17$, $p < .05$). The results are consistent with previous analyses and demonstrate that the overall recognition performance of analytics is consistently superior to that of globals. The results also indicate that the recognition performance of both groups is enhanced by providing additional memory cues, and that the facilitiation due to cueing is greater for analytics than for global individuals. The results also suggest that memory ability could also be playing a part in the discrimination differences exhibited between analytics and globals.

EXPERIMENT 2

One goal of Experiment 2 was to replicate the analytic versus global cognitive style descrimination findings while eliminating the memory component from the task. In addition, two new cognitive style measures were investigated and two factors related to the discrimination task were manipulated.

Method

Subjects - Twenty student volunteers from a local technical college served as subjects.

Procedures - The procedures of Experiment 2 were similar to those of Experiment 1 with the following changes and alterations. Two additional style variables were assessed, anxiousness and risk taking. The induction materials were intensified to: (1) present a more engaging cover story and (2) detail a rationale for the sources of uncertainty.

The number of launch sites was increased from 10 to 12, and associated with each site was a probability indicating the likelihood that an attack would be launched from that site. A random half of the probabilities indicated a low chance of attack (.15 to .25) while the other half indicated a high chance of attack (.45 to .60). High and low probability sites were counterbalanced geographically and the probabilities were clearly visible on all maps.

At variance with Experiment 1, the attack and test trajectories were drawn on each map making it unnecessary for subjects to memorize them. The heat sensor readings were also manipulated. For a random third of the events,

the heat sensor readings were clearly out of range (i.e., 1, 2, 11, 12) for another third of the events, the readings were within the error range (i.e., 3, 4, 9, 10), and the for the remaining third, the readings were within the known heat signature of the missiles (i.e., 5, 6, 7, 8). When plotting the position of the heat source, an error radius of from 0 to 20 miles was introduced.

As before, subjects rated the likelihood of an attack. However, this time a 9 point scale was used, where 1 meant a 10 percent chance of attack and 9 meant a 90 percent chance of attack. Subjects were given several practice trials, and then 36 events of three reports each were conducted.

Results and Discussion

Analyses of the style measures showed that, once again, analytics discriminated better between attack and test conditions than global subjects (respective means for high and low analytic subject were for d', 1.31 and 0.83, and for area under ROC curve, 89.93 and 71.36; in both cases, the differences are significant at $p < .02$). Thus, the discrimination differences obtained in Experiment 1 were not the results of memory ability differences alone. Similar analyses for anxiety and risk taking revealed neither to be related to discrimination ability.

To examine the manipulations of the various sources of uncertainty on multiple cue decision making, the attack probability ratings were analyzed in an outcome (attack/test) × site probability (high/low) × heat signature range (out/error/in) × reports (1/2/3), four-way within-subjects ANOVA. All four main effects were significant. Subjects recorded higher attack probabilities: (1) in attack conditions, (2) when site probability was high, (3) as the heat signature range moved closer to in-missile range, and (4) with each succeeding report (all p's $< .002$). All the two-way interactions also proved significant, and their results are summarized in Tables 1a through 1f. The outcome × site probability × reports and the outcome × heat signature range × reports interactions were also significant, but these three-way interactions will not be discussed here.

Referring to Table 1a, we see that prior probability of missile site has a greater impact in the test than in the attack situation. Table 1b shows that subjects see attack likelihood more similar in attack and test situations when heat signature is clearly in the missile range, than in either of the other two conditions. Inspection of Table 1c reveals that subjects appropriately increment their attack ratings in attack situations and decrement their attack ratings in test situations with each new report update. From Table 1d, we see that heat signature information has little effect on attack ratings when site probability is high, but when site probability is low, attack likelihood increases as heat signature approach in-missile range. The data in Table 1e show a similar result. The additional information in each report has no affect when site probability is high; however, when site probability is low, a second report increments attack probability considerably. These latter two findings, together with the first, show the pervasive effect of prior site attack probability data on attack ratings. Table 1f shows that second report information has an effect only when heat signature is in the error or missile range. The third report produces a small increase in attack probability when heat signature is out of range, a moderate increase when it is in the error range, and a moderate decrease when it is in-missile range. Subjects seem to have some difficulty integrating additional information when heat signature is in-missile range.

In conclusion, the signal detection analyses of both experiments indicates that subjects differing in cognitive style also differ in the way they make discriminations. Analytic subjects clearly discriminated better between attack and test situations than global subjects and differences in memory ability between the two groups cannot account for this difference in discrimination ability. The recognition memory analysis showed that analytics also made superior use of incremental cue information present in each report update. The ANOVA results demonstrated that prior information, such as the likelihood that an attack would be launched from a particular site, bias the way subjects gauged the probability of attack. Missile site attack probability is apparently seen as quite concrete, subjects anchor on it, and weigh this source of information more than report updates and heat signature. Heat signature and report update information prove more difficult for subjects to use. Heat signature data seemed most beneficial to making discriminations when readings were clearly out of range, in the error range, or when site probability was low. Information provided by report updates had its largest impact when other sources of information were weak.

REFERENCES

Benbasat, I. and A. Dexter, "Value and Events Approaches to Accounting: An Experimental Evaluation," Accounting Review, October 1979.

Green, D.M. and J.A. Swets, Signal Detection and Psychophysics, New York: Wiley, 1966.

Witkin, A.A., P.K. Oltman, E. Raskin, and S.A. Karp, A Manual for the Group Embedded Figures Test, Palo Alto, California: Consulting Psychologists Press, Inc., 1971.

TABLE 1. TWO-WAY INTERACTIONS FOR ATTACK PROBABILITY

a) OUTCOME × SITE PROB.
$F = 7.12$, $df = 1,19$, $p < .02$

	SITE PROB	
OUTCOME	LOW	HIGH
ATTACK	5.13	5.80
TEST	2.99	4.02

b) OUTCOME × HEAT SIGN.
$F = 8.77$, $df = 2,38$, $p < .001$

	HEAT SIGN. RANGE		
OUTCOME	OUT	ERROR	IN
ATTACK	5.28	5.63	5.48
TEST	3.22	3.37	3.94

c) OUTCOME × REPORTS
$F = 85.78$, $df = 2,38$, $p < .001$

	REPORTS		
OUTCOME	FIRST	SECOND	THIRD
ATTACK	4.33	5.44	6.62
TEST	4.10	3.72	2.71

d) SITE PROB. × HEAT SIGN.
$F = 15.30$, $df = 2,38$, $p < .001$

	HEAT SIGN. RANGE		
SITE PROB.	OUT	ERROR	IN
HIGH	4.99	4.75	5.00
LOW	3.52	4.24	4.42

e) SITE PROB. × REPORTS
$F = 9.71$, $df = 2,38$, $p < .001$

	REPORTS		
SITE PROB.	FIRST	SECOND	THIRD
HIGH	4.86	4.90	4.98
LOW	3.57	4.27	4.35

f) HEAT SIGN. × REPORTS
$F = 11.65$, $df = 4,76$, $p < .001$

	REPORTS		
HEAT SIGN.	FIRST	SECOND	THIRD
IN	4.42	5.02	4.70
ERROR	4.07	4.53	4.89
OUT	4.15	4.20	4.41

Cognitive Processing Capabilities Relevant to Design of Visual Displays in Advanced Airborne Weapon Systems

Dr. Wade R. Helm and Dr. Tommy R. Morrison

Naval Aerospace Medical Research Laboratory
Naval Air Station, Pensacola, Florida 32508-5700

Abstract

Critical design considerations for information management and decision making in advanced airborne weapon systems depend upon understanding the demands imposed upon the human operator's cognitive processing capabilities. An integrated design approach for optimal utilization of advanced displays, decision aiding support systems, and artificial intelligence-based systems must address the issue of task(s) demands imposed upon the operator and defined in terms of the operator's cognitive capabilities for processing multiple types of information. Potential gains in airborne system performance, made possible by the incorporation of these technologies, may be offset by failure to adequately incorporate human cognitive capabilities relevant to complex tactical display situations during the design of these multi-technology integration display systems.

A function-analysis for a Navy Tactical Coordinator aircrew position was utilized as the basis for the design of a tactical visual display task employed in the present investigation. This task required subjects to selectively attend to targets specified in terms of the following: (1) shape, (2) size, (3) color, (4) heading, (5) position, and (6) quantity. Task difficulty was manipulated by: (1) varying the number of relevant target characteristics; (2) requiring certain targets to be mentally rotated; and (3) requiring performance of a mental arithmetic operation on particular targets. Certain combinations of the relevant target information had to be retained in memory while the tactical display task was visually searched for the specified targets.

Perceptual processing times and accuracies were measured and examined as a function of the amount and type of relevant target information to which subjects were instructed to attend. Separate stepwise multiple-regression procedures were used to analyze the times to encode the task information and the times to perform the subsequently presented tactical display task. Regression analysis of the encoding times provided evidence for 10 cognitive processes (R^2 = .80, N = 152). Regression analysis of display processing times showed evidence of 8 cognitive processes (R^2 = .60, N = 152). The data obtained in the study provide understanding of cognitive capabilities, which may lead to effective integration of humans into emerging advanced tactical information management and decision making systems.

Disclaimer Statement

Opinions or conclusions contained in this report are those of the authors and do not necessarily reflect the views or endorsement of the Navy Department.

The research reported in this paper was completed under the Naval Air System Command work unit 61153N WR04210001.6142

Introduction

The magnitude of threats in today's aviation environment have necessitated the development of highly sophisticated hardware/software systems to perform additional sensor, weapons, and command-control functions. As a result, aviation weapon system operators are often overburdened with information processing and output control requirements to the extent that task performance can deteriorate and thereby compromise mission effectiveness and safety. For example, previous tactical air missions depended predominantly upon a pilot's psychomotor abilities in guiding his aircraft and its projectiles, but in today's complex aviation system the pilot relies also upon his ability to process vast arrays of complex information, to make rapid, highly-consequential decisions, and execute a large number of coordinated responses. New technologies such as Voice Interactive System (VIS) and computerized Decision Support Systems (DSS) represent considerable potential for alleviating some of the processing burden imposed upon operators of modern systems. In order to integrate effectively VIS or DSS in future system designs, however, we must be able to predict the conditions under which an aviator's sensory-perceptual-cognitive capabilities are exceeded by situational information-processing demands. Furthermore, we must be able to verify, through appropriate tradeoff evaluations, the realistic utility of these and other advanced man-machine interfacing techniques for enhancing performance relative to their potential for further aggravating the demand problem. Therefore, improved methods must be developed to allow assessments of the influence of system and task demands on the limited processing resources available to the human operator and to provide an analytic framework from which prediction can be made about performance in novel situations.

The primary objective of the current Naval Aerospace Medical Research Laboratory (NAMRL) program is to provide meaningful, performance based definitions of operator cognitive capabilities and limitations across broad categories of aviation-relevant tasking and workload requirements. Specific objectives include: (1) the development of a tasking and measurement system for defining human cognitive capabilities and limitations with respect to aviation requirements; (2) the development of a process-based model of cognitive capabilities for predicting performance in the aviation environment; (3) the refinement of techniques to define and quantify operator workload capabilities in terms of mission/system demands; and (4) the development of a user-oriented human factors data base.

The purpose of this paper is to present the results of a series of experiments designed to meet program objectives of identifying and defining operator cognitive processing capabilities and limitations in aviation relevant tasks. The present investigations employed a function-analysis for a Navy Tactical Coordinator aircrew position as the basis for the design of a tactical visual display task. This task required subjects to selectively attend to targets specified in terms of certain characteristics. Relevant target information had to be retained in memory while the tactical display task was visually searched for the specified targets. It was hypothesized that processing times and accuracies would be a function of the amount and type of relevant target information to which subjects were instructed to attend. Analysis of times and accuracies data should reveal specific underlying cognitive processes required in complex aviation, information management and decision making tasks.

Method

Subjects. Nine Naval and 43 Marine Officers entering the navy flight program at Pensacola, Florida participated in this study. All 52 subjects were male. All had 20/20 or better visual acuity and screened for color vision with the Armed Forces Vision Tester. Subjects' ages ranged from 22 to 30 years, with a mean age of 24 years.

Apparatus. The test station consisted of a test booth enclosure in which the seated S performed the experimental task. A television monitor and Caramate rear-projection slide system were positioned in front of the S, with the monitor left of center and the projector right. A response keypad was positioned directly in front of the S. The keypad contained keys labeled 0-9, True, False, and Enter. An Apple microcomputer interfaced to an MCDS switching system controlled task presentation and recorded S response times in msec.

One hundred and twenty slides were presented on the display screen. The illuminated display screen area was 15.25 cm x 15.25 cm and divided into quadrants by horizontal and vertical lines (each had 1 mm stroke width). A 7.62 cm diameter circle (stroke width = 1 mm) was centered in the display screen. Objects presented on the display screen varied in shape (triangle, rectangle, pentagon), color (red, green, white), size (small, medium, large), heading (N, NE, E, SE, S, SW, W, NW), and screen location (random). Number of symbols per slide varied from 10 to 20. The dimensions of the symbol sizes measured on the surface of the display screen are presented in Table 1. Each symbol enclosed a solid black triangle which indicated symbol heading. In 60 slides all symbols were the same color; in 13 and 47 slides, symbols were of two or three colors, respectively. For all cases, each symbol was one color. Triangles represented airplanes, rectangles and pentagons represented aircraft carriers and destroyers, respectively. Thus, e.g., carriers were of either three sizes and three colors.

Various questions were presented on the TV monitor in all capital letters. Questions differed by amount and type of information asked; for example, "How many red carriers are on the screen?". An example of a difficult question was "At least 2 small red destroyers heading south are in the upper screen portion (True or False)?". The simple question required subjects to memorize and recall two types of information: color = red; and shape = carrier to successfully respond to the subsequently presented display screen. The difficult question included the following kinds of information to be memorized and recalled: (1) number of question objects = 2; (2) size = small; (3) color = red; (4) shape = destroyer; (5) heading = South; and (6) screen portion = upper. The questions were formed by combining the question elements presented in Table 2.

Procedures

Taped verbal instructions (approximately 5 minutes) with programmed example slides were presented via projector to each subject seated at the test station. Following instructions, the experiment began. The experiment consisted of the presentation of three slide groups (A, B, and C), each containing 40 slides. The order of slide presentation within slide group was constant, however, order of slide group presentation was random across subjects. The experimenter started the first trial of each slide group and thereafter the experiment was self-paced. A trial consisted of the following: (1) a question appeared on the TV monitor, (2) subject read the question, (3) subject pressed "Enter" (reaction time 1) which resulted in the simultaneous removal of the question from the TV monitor and

presentation of the display slide, (4) subject visually examined the display slide and responded via keypad (reaction time 2) in accordance with the immediately preceding question, (5) display slide was removed from view and (6) feedback was presented on the TV monitor. When subject pressed "Enter" again the next trial began. Reaction times 1 and 2 (RT 1 and 2) were measured in msec. Subjects required 20-25 minutes to complete each slide group.

Analysis. Means were computed across subjects for each slide as shown in Table 3 for RT1, RT2, and Accuracy Scores. Mean RT1 and Mean RT2 data were the dependent measures used in separate step-wise regression analyses. As shown in Table 3, the questions were coded in accordance with their content elements (previously defined in Table 2) and the coded question element data were independent variables for the regression analyses. As indicated in Table 3, it is assumed that certain required cognitive processes are **postulated** in order to process the question element information and execute the stored information in the display context. In Table 3, certain variables were coded either "1" or "0." The 1 indicated that the question contained that element. The following three variables were continuous: Number of Screen Objects, Number of Targets, and Number of Question Objects, and were coded appropriately. For example, the question for slide 1 was: "Exactly five objects are in the right portion of the screen (True or False)?" Thus the relevant question elements for RT1 regression analysis were: Screen Portion (dummy code = 1), and Number of Question Objects = 5.

For the RT2 regression analysis Number of Screen Objects and Number of Targets variables (and values) were available for entry into the regression model. Number of Screen Objects refers to the total number of symbols on a display slide. The Number of Targets refers to the number of displayed symbols which fit the question. Since the Number of Screen objects and Number of Targets information could not affect reading the question (RT1), these latter two variables were available for entry only in the regression analyses for RT2.

Results

The results of the stepwise regression analyses for RT1 (Encoding Reaction Times) are presented in Table 4. This accounts for 81% of the variance in times required to encode the question information. The abbreviated variables in Table 4 are defined as follows:

SP1 = Screen portion; CP1 = Circle position; CM1 = Circle movement; SHP1 = 1 shape; H1 = 1 heading; SZ1 = 1 size; RED2 = All red objects represent 2 objects; ROT = Mentally rotated red objects; NQO = Number of question objects; SZ2 = 2 sizes; CLR2 = 2 colors.

Thus, 10 question element variables accounted for significant portions of the total RT1 variance. Comparisons can be made between question elements B-values in order to compare required processing times for the variables of interest. For example, the B-value for ROT (mental rotation) required the most time (5.11 sec) while relatively little time was needed to encode question content for 1 shape (0.92 sec).

The results of the stepwise regression analysis for RT2 (display reaction times) are presented in Table 5. Eleven independent variables made significant contribution to the times required to process the visual displays; total variance accounted for 67%. Abbreviated variables in Table 5 which were not defined immediately above are:

CLR1 = 1 color; NTGT = Number of targets; NTOT = Number of total symbols; SHP2 = 2 shapes; H2 = 2 headings.

Again, comparisons can be made among B-values for the independent variables of interest. The B-values can be interpreted as the time to perform the cognitive process needed to process the particular independent variable demand.

Discussion

The methodology presented above provides a capability to assess cognitive processing demands, in terms of associated processing times, for particular critical task variables of interest within the context of a complex visual display. It is noteworthy, that certain variables required significantly large encoding and display reaction times (e.g., mental rotation) and such demands should be avoided in display-task design. Certain variables (SP1, CP1) which reduce the display area to be visually searched obtained negative B-values, thus, indicating a beneficial effect in display processing times. It is evident that these display-task features should be included, where possible, in display design. Where a single target color differed from the other display symbols the B-value obtained a negative value (CLR1 = -1.41 sec) which is congruent with the literature concerned with effects of color on visual search. Further, NTOT and NTGT contributed positively to increasing the required display processing times, as would be expected from known effects of display density on visual search. Ongoing data collection and analysis (e.g., hierarchical factor analysis) at NAMRL should provide analytic evidence as to relative importance and statistical distribution of the identified cognitive processes. This analysis should provide critical information as to implication for interface design of future aviation weapons systems.

TABLE 1

Symbols and projected dimensions (mm)

Symbol		Small	Medium	Large
	H =	10	13	16
	L =	7	9	10
	H =	9	11	17
	L =	6	7	10
	H =	10	13	17
	L =	7	8	10

TABLE 2

Question Elements

Circle Position	Targets were designated as either inside or outside the circle.
Screen Portion	Targets were designated as: Upper, Lower, Right, Left, Upper-Left, Upper-Right, Lower-Left, Lower-Right. No specified screen area meant targets were in the full screen.
Circle Movement	Targets were designated as moving: To or away from the center of the circle; or would pass through or enter the circle.
Shape	*Triangles = airplane; rectangles = carrier; pentagons =* destroyer.
Color	Red, green, white.
Size	Small, medium, large.
Heading	North, East, South, West, NE, SE, SW, NW.
All red objects represent 2 objects (this required subject to count each red object as really representing 2 objects).	
Assume all red objects are rotated 90 degrees to the right (this required subjects to spatially rotate certain display symbols).	
Display Screen Objects	Total number of symbols on display screen.
Target Objects	Number of target objects to be searched for in display screen.
Number Question Objects	Number of target objects specified in question.
2 Shapes	Certain questions included 2 shapes; e.g., carriers and airplanes.
2 Headings	Certain questions called for target objects heading in either of 2 specified headings.
2 Sizes	Certain questions called for targets of 2 sizes.
2 Colors	Certain questions called for targets which were either of 2 colors.

A Preliminary Investigation of Problem Solving and Judgmental Strategies of Expert Military Intelligence Personnel

Vanessa Irizarry, Ph.D., & Beverly Knapp, Ph.D.
United States Army Research Institute

Abstract

This research was part of a larger effort aimed at identifying the cognitive processes that underlie the production of military intelligence (MI). Eight tasks were administered to 113 MI experts in order to determine any cognitive skills or information processing strategies associated with expertise in the MI domain. A number of strategies were identified in terms of four major aspects of problem solving: hypothesis testing, pattern identification, integration of probabilistic information, and problem structuring.

Background

Military intelligence analysis consists of a number of complex problem solving activities. The solutions to these problems are frequently articulated in the form of decisions (e.g., targeting, collection management, selection of probable courses of action, etc.). Decisionmaking is therefore an integral part of the problem solving that underlies intelligence analysis. MI decisionmaking is not characterized by clear-cut choices among well defined alternatives. Prior to actual decisionmaking, analysts must process and make judgments based on large amounts of data which are frequently incomplete and of questionable validity and reliability (Thompson, Hopf-Weichel & Geiselman, 1984). The present research is aimed at understanding these "predecisional" cognitive processes within the MI context.

Predecisional behavior refers to the cognitive processes whereby a problem is recognized and structured and its informational inputs are apprehended, evaluated, interpreted and articulated in the form of alternative solutions. A considerable amount of psychological research has shown that systematic judgmental errors or "cognitive biases" affect decisionmaking and, in particular, the judgmental operations that precede it (Kahneman, Slovic & Tversky, 1982). The present paper focuses on predecisional judgmental operations and on the cognitive biases known to affect them.

Objectives

An understanding of the cognitive bases of MI is essential for the development of effective automated support and training strategies for new and experienced intelligence analysts. The present report explores the extent to which cognitive biases affect the predecisional behavior of expert MI personnel. These findings will contribute to the development of a cognitive profile of MI expertise which specifies the cognitive skills and strategies as well as any biases that may characterize experienced MI personnel. This profile can then be used to generate training recommendations for new personnel and specifications for automated support to avoid or minimize cognitive biases.

Procedure

Eight paper-and-pencil tasks to investigate problem solving and judgment under uncertainty were extracted from relevant basic research in cognitive psychology. Whenever possible, these tasks were modified to include content and examples derived from operational MI settings. This was done in order to enhance as well as to test the external validity of these tasks to the MI domain. The tasks addressed four major aspects of problem solving and judgment: (1) testing hypotheses, (2) identifying patterns of covariation among events, (3) integrating new and old probabilistic information, (4) problem structuring strategies used to handle large data sets. The eight tasks were incorporated in a booklet which was administered to a sample of 113 expert MI personnel.

Results and Discussion

(A) Hypothesis Testing

During hypothesis testing, people adopt one of two major strategies, or a combination of both. A confirmatory set refers to the tendency to select evidence that conforms to a given hypothesis whereas a disconfirmatory set refers to the tendency to seek evidence that can potentially falsify it. Two hypothesis testing tasks which have been widely used in previous research were included: (1) Wason's "four card selection task" (Wason, 1968), and (2) a version of the "2-4-6" rule discovery" task (Wason, 1960) adapted for group testing. An abstract and an "applied" version (incorporating MI thematic content) of the four-card task were administered. For all tasks, expert MI personnel exhibited a strong tendency to use confirmatory over disconfirmatory strategies. Although the modal strategy for the applied task was confirmatory, the frequency of confirmatory choices decreased and that of disconfirmatory choices increased for the applied four-card task, relative to its abstract version. These results indicate that the ability to appreciate the value of and to use

disconfirmatory strategies is enhanced by presenting the problem within more realistic, operational MI contexts.

In sum, the observed prevalence of confirmatory hypothesis testing strategies among MI experts in both abstract and applied contexts points to the need to develop automated support and/or training methods to enhance the use of disconfirmatory strategies. However, the observed increase of disconfirmation in applied tasks indicates that further research should investigate hypothesis testing in more realistic intelligence simulations.

(B) Pattern Identification

The ability to recognize functional relationships (or the absence of the same) among battlefield events is central for the identification of patterns of behavior which, in turn, serve to describe and predict enemy activities. The present task investigated strategies used to judge the relation between battlefield events based on the frequencies with which they co-occur. Results indicated that unlike most people (see Ward & Jenkins, 1965), MI experts are quite successful at making judgments about the presence/absence of functional relationships based on event frequencies. This skill was exhibited in two ways: (1) MI experts take into account all the relevant information needed to make accurate judgments, (2) a smaller proportion also take into account the impact of the sample size of observations for a given event on the reliability of the information that the event provides. Since experts were highly proficient in this task, additional research should be conducted with less experienced, novice personnel to determine their level of proficiency in pattern identification.

(C) Strategies for integrating new and old probabilistic information

The production of intelligence products frequently involves the processing of uncertain or probabilistic information. From a formal logical and statistical perspective, judgments about the probability of a given target event necessitate the consideration of not only the specific evidence about the case at hand but also the previously observed or "base-rate" frequency of the target outcome in some relevant reference population (Tversky & Kahneman, 1982). Two scenario-based problems were designed to look at (1) the factors that MI experts take into account when judging the informative value of recent data (intelligence messages), and (2) strategies that they use to combine old and new probabilistic information. Three major findings obtained: (1) most subjects adjusted the probabilities of the recent messages about the occurrence of a target event according to previous base-rate probabilities; (2) most subjects did not use a strictly Bayesian approach to integrate previous and recent data,

although a majority used mathematical methods to generate their judgments; (3) a minority totally ignored the impact of previous probabilities on their judgments about the informative value of the recent messages. Further research should investigate whether novice MI personnel employ similar strategies. In addition, the non-Bayesian methods that MI experts employ to integrate probabilistic information need to be described in more detail.

(D) Problem structuring

The way in which a problem is structured has been shown to affect the efficiency with which it is solved (Schwartz, 1971). This is probably due to the fact that certain problem structuring strategies permit a more clear assessment of the information required for its solution. Five distinct strategies were identified on the basis of written protocols generated by each subject during problem solving each of two versions of a deductive reasoning task. One version provided all the information required to draw a valid deductive inference whereas the other version contained only partial information so that subjects had to identify the missing data required to draw a valid deduction. For both problem conditions, the frequencies and solution rates associated with each strategy indicated that a "matrix" approach was the most successful, although very few subjects employed this strategy. The patterning of the data across problem conditions revealed an "information seeking bias". This tendency to seek additional information could have accounted for the fact that performance on the partial information condition was better than that on the complete information condition. These data suggest that the information for intelligence tasks requiring the integration of a large number of messages along various dimensions should be presented in a matrix-like format which is compatible with the most successful strategy observed. These data also indicate the need to explore the trainability of problem structuring strategies.

In conclusion, this work identified a number of cognitive skills and biases characteristic of MI experts. The findings have important implications for both training and automated systems design. However, a number of issues need to be resolved before these implications can be translated into formal recommendations. As noted in the text, research with less experienced novice personnel is needed to establish the uniqueness of these findings to MI experts. In addition, further research should investigate the relationships between each one of the cognitive skills described above and on-the-job performance measures to determine whether performance in these cognitive tasks can be used to predict performance in operational MI settings.

References

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). Judgment under uncertainty: Heuristics and biases. New York, NY: Cambridge University Press.

Schwartz, S. H. (1971). Modes of representation and problem solving: Well evolved in half solved. Journal of Experimental Psychology, 91(2), 347-350.

Thompson, J. R., Hopf-Weichel, R., & Geiselman, R. E. (1984). The cognitive bases of intelligence analysis. ARI Research Report No. 1362 (AD-A146 132).

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. New York, NY: Cambridge University Press.

Ward, W. C., & Jenkins, H. M. (1965). The display of information and the judgment of contingency. Canadian Journal of Psychology, 19(3), 231-241.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. Quarterly Journal of Experimental Psychology, 12, 129-140.

Wason, P. C. (1968). Reasoning about a rule. Quarterly Journal of Experimental Psycholgy, 20, 273-281.

ACKNOWLEDGEMENTS

This research effort would not have been possible without the cooperation of the U.S. Army Intelligence Center and School (USAICS). Appreciation is extended to Dr. Julie Hopson and LTC Miles Kara for coordinating ARI's participation in the G2/Commanders Military Intelligence Conference, and to CPT Michael Mather for his technical assistance.

A more detailed ARI report can be obtained from Dr. Irizarry at the U.S. Army Research Institute, PERI-SF, 5001 Eisenhower Ave., Alexandria, VA 22333-5600.

PANEL SESSION

Title: STRESS AND PERFORMANCE

Chair: Gerald A. Hudgens, Ph.D. (US Army Human Engineering Laboratory)

Discussant: James P. Torre, Jr. (US Army Human Engineering Laboratory)

SLEEP AND FATIGUE IN OPERATING FLIGHT CREWS

R. Curtis Graeber, Ph.D. (NASA, Ames Research Center)

PHYSIOLOGICAL AND BEHAVIORAL RESPONSES TO THE STRESS OF SLEEP LOSS

Laverne C. Johnson, Ph.D. (Naval Health Research Center)

PROBLEMS OF USING SLEEP RESEARCH TECHNIQUES IN THE FIELD

Gerald P. Krueger, Ph.D. (Walter Reed Army Institute of Research)

THE ASSESSMENT OF WORKLOAD EFFECTS

Glenn F. Wilson, Ph.D. (Armstrong Aerospace Medical Research Laboratory)

SUBJECTIVE SYMPTOMS, CREW PERFORMANCE, AND COGNITIVE INTERVENTIONS

Terry Rauch, Ph.D. (Army Research Institute of Environmental Medicine)

PROBLEMS IN MODELING COMBAT STRESS: A PROGRAM TO MEET THE CHALLENGE

Gerald A. Hudgens, Ph.D. (US Army Human Engineering Laboratory)

Sleep and Fatigue in Operating Flight Crews

R. Curtis Graeber
NASA Ames Research Center

Abstract

The effects of flight duty on aircrew sleep and fatigue were examined in commercial and military crews flying short- and long-haul routes. Sleep duration and quality are adversely affected in short-haul operations despite mostly day and evening flights crossing a maximum of one time zone. These effects are reflected in subjective mood and fatigue changes during the trip. Sleep disturbances are more prevalent in international long-haul crews due to circadian desynchronization. Polysomnographic recordings coupled with daytime multiple sleep latency tests have confirmed that greater disruption of sleep-wake patterns occurs after eastward than after westward flights. Factors such as age, personality, and exercise are important determinants of the extent of these effects in individual crew members.

Introduction

Recent technical advances in aircraft design and air traffic control have resulted in a substantial improvement in aviation safety; however, a persistent 65 percent of all mishaps in commercial aviation continues to result from flight crew error. A variety of human factors underlie these statistics, but fatigue has always been one which has attracted attention. To minimize its contribution, both civilian and military authorities have placed limits on flight time and duty periods. Regardless, the complexity of today's flight schedules and crew composition suggests that other approaches are needed to help individual crew members and operators develop the safest, most efficient interface between operational demands and the sleep-wake characteristics of the human operator.

Consequently, we have undertaken a series of studies to determine the impact of duty schedules on the sleep and circadian rhythms of operating flight crews and the corresponding variations in fatigue and mood. At the same time we are attempting to identify those individual attributes which may determine, or help predict, the responses of crew members to various schedules. A third focus is on the development of strategies or countermeasures which individual crew members or operational managers can employ to cope with such effects and thus minimize their impact. Finally, the operational significance of these effects is being evaluated using full-mission simulation scenarios. This paper will concentrate on the first two topics, i.e., describing the psychophysiological effects of flight duty and the role of individual attributes; other references are available which more fully describe the other aspects of the overall research program (Graeber, Foushee, & Lauber, 1984; Graeber, Foushee, Gander, & Noga, 1985; Foushee, Lauber, Baetge, & Acomb, 1986).

Method

Two basic study designs have been used. The first requires continuous (i.e., 24h/day) physiological monitoring of crews before, during, and after a regularly scheduled trip. Heartrate,

rectal temperature, and non-dominant wrist activity are recorded every 2 min. by an ambulatory Vitalog PMS-8 solid state monitor (Vitalog Corp., Redwood City, CA). Subjects also maintain a pocket-size daily log in which they enter a variety of data concerning sleep times and quality, exercise, meals, voidings, bathing, symptoms, and duty. Mood and fatigue self-assessments are recorded every even GMT hour. Volunteers also complete a background questionnaire which provides information on individual attributes , e.g., lifestyle, personality, sleep habits, etc. A trained NASA cockpit observer accompanies the crew on the flight deck during all flights and remains with them throughout the trip to record operational data and reinforce subject cooperation. Depending on the exact trip, these recordings can begin up to three days before departure, continue for three to eight days of duty, and terminate up to five days later.

The second approach has been used less extensively, but provides a more in-depth analysis of the sleep disturbance experienced by international flight crews. Through the cooperation of both U.S. and foreign international carriers and their pilot associations, crew members spent their first layover after departure from homebase in a sleep laboratory instead of the usual hotel. Prior to this time they had undergone nocturnal polysomnographic recordings which included EEG, EOG, and EMG at a minimum, plus respiration, ECG, and leg movements for most subjects. All participants underwent one adaptation night and, if possible, one baseline night of recordings before departure. Some baseline recordings had to take place several days after the trip due to scheduling difficulties. In addition, multiple sleep latency tests (MSLTs) were administered every even GMT hour after the baseline recording (i.e., 12 h) and throughout the layover to obtain objective measures of daytime sleepiness. During the layover (usually 48h) crew members were instructed to follow their usual sleep-wake pattern for that destination. They could sleep or eat whenever they chose, mild exercise was readily available, and MSLTs were administered whenever they were awake. Continuous recording of body temperature was offered as an individual option to those who wished to provide supplementary circadian data. A total of four B-747 crews participated on each of five flight routes all requiring time zone transitions of 7 to 8 hours: Frankfurt, London, or Tokyo to San Francisco, and San Francisco to London or Tokyo. Sleep recordings were conducted at the Stanford Sleep Research Center, the DFVLR Institute for Aerospace Medicine, the RAF Institute of Aviation Medicine, and at various hotels in Tokyo.

Results

The primary data for the in-flight field study comes from two-years of data collection on 91 pilots on 47 short-haul commercial trips requiring 821 flights and 924 flight hours. All flights were in DC-9, B-737, or B-727 aircraft and were concentrated in the high density air space of the eastern U.S. This sample represents 84.7% of all crewmembers contacted and asked to participate in the study. Volunteers were solicited from crews who had already been assigned to the trip, and the confidentiality of all data was guaranteed by the use of a blind coding scheme. The results presented here come from 74 of these subjects, excluding the other 17 crew members who flew B-727 transcontinental night flights. Similar data has also been collected on USAF MAC C-9 crews flying routes comparable to the commercial DC-9 crews; on MAC C-130 crews deploying from Dyes AFB, TX, to Mildenhall AFB, U.K., and on Royal Norwegian Air Force P-3 crews flying from Andoya Island, north of the Arctic Circle, to Moffett Field, CA, for one week of simulator training before returning.

Short-haul aircrew sleep is significantly shorter (1h) on trips than at home. This effect is manifested in progressively shorter sleep durations over successive layover nights. The decrease in duration is accompanied by significantly reduced sleep quality and an increased number of nocturnal awakenings. Although overall mean heart rate and activity during sleep were not

significantly different from baseline, high levels of heart rate and activity during sleep were significantly correlated with poorer sleep ratings and shorter sleep durations. A high frequency of napping on the first day post-trip, together with shortened nocturnal sleep latencies post-trip, suggests the recuperation of sleep debt accumulated on trips. The impact of poorer sleep is reflected in a significant increase in subjective fatigue and more negative moods over successive trip days.

Key operational factors in the observed sleep changes were early on-duty report times and the number of segments flown per day. Length of duty did not emerge as an important independent factor. Also, length of the trip (i.e., 3 vs. 4 days) did not significantly affect subjective fatigue or mood; however, time-of-day was found to be an important variable modulating these psychological effects.

Individual crew member attributes, as determined from the background questionnaire and daily log, were surprisingly powerful in accounting for much of the variation in sleep and fatigue. The primary effect was seen in the correlation of certain personality indices with subjective sleep quality, heart rate during sleep, and fatigue and mood ratings during the trip. Those crew members who scored on higher on measures of "instrumentality " and "expressiveness" reported sleeping better and less on trips than those who scored lower on those dimensions. Age (25-58y) also appears to be a significant factor in short-haul flight operations, with older subjects tending to manifest greater increases in heart rate and activity during sleep on layovers and reporting significantly higher levels of fatigue at the end of the duty day than younger crew members. Caffeine and alcohol consumption increased significantly on trips, but only alcohol detrimentally affected sleep quality. The frequency of exercise decreased on trips, but those subjects who exercised both at home and on trips manifested the smallest changes in heart rate during sleep on trips as compared to home.

As expected, the international air crews exhibited substantially more sleep disruption during their layovers than the short-haul crew members. The polysomnographic recordings revealed that sleep architecture was significantly altered after multiple time zone flights, with the type of alteration (e.g., decreased REM sleep) depending on the direction of flight . Overall, the amount of sleep disturbance was not as severe as expected, and most crew members achieved adequate sleep by either sleeping quite well at a particular time of the 24h day or by staying in bed longer and sleeping in several bouts. Eastward flights produced greater disturbance in sleep-wake patterns than did westward flights. The impact of this directional asymmetry may be lessened if crew members adopt more stringent sleep discipline and limit their sleep upon arrival to about two hours or less. Individual variation was substantial, with age being significantly correlated with poorer sleep quality and increased daytime sleepiness.

Discussion

These results demonstrate that flight crews often experience sleep difficulties even when duty schedules are within legal limits and rest periods appear to be adequate. Although these difficulties are reflected in subjective fatigue and mood during duty, most crews are able to compensate for these effects during the length of the trips (see Foushee et al., 1986). One factor which has not been carefully examined yet is the amount of time needed for recovery between trips. There is little doubt that cumulative sleep deficits do occur in both short-haul and long-haul flying and that crew members may require more that one or two days to recover. The latter requirement becomes greater when more time zones have been crossed and the duration of overseas layovers is

increased.

The more unique aspect of these findings is that they confirm the usefulness of individual strategies to reduce the impact of such flight schedules and reveal the potential importance of individual attributes for selecting crew members who may be less affected by these schedules. Both of these approaches differ considerably from the traditional regulatory approach which proscribes flight and duty time limitations to address the problem of flight duty and fatigue. While the latter is obviously still an effective way of placing outer limits on the potential for excessive crew fatigue, the opportunity now exists for developing a safer, healthier, and more efficient interface between operational demands and the sleep-wake characteristics of aircrew.

References

Foushee, H.C., Lauber, J.K., Baetge, M.M. & Acomb, D.B. (1986). Pre- and post-trip operational performance in short-haul flight operations: a full-mission simulator study. In Proc. 38th Annual International Air Safety Seminar. Arlington: Flight Safety Foundation, in press.

Graeber, R.C., Foushee, H.C., & Lauber, J.K. (1984). Dimensions of flight crew performance decrements: methodological implications for field research. In H.M. Wegmann (Ed.), Breakdown in human adaptation to 'stress', (vol. 2, pp. 584-605). The Hague: Martinus Nijhoff.

Graeber, R.C., Foushee, H.C., Gander, P.H., & Noga, G.W. (1985) Circadian rhythmicity and fatigue in flight operations. In Proc. 4th UOEH Int'l. Symp. on Occupational Health in Aviation and Space Work. J. of UOEH: 7 (suppl.): 122-129.

PHYSIOLOGICAL AND BEHAVIORAL RESPONSES TO THE STRESS OF SLEEP LOSS

Laverne C. Johnson, Ph.D.
Naval Health Research Center
San Diego, California

Abstract

Total sleep loss of greater than 60 hours may be expected to produce some physiological, biochemical, performance, behavioral, and mood changes. The degree of change depends upon the individual but, as sleep loss progresses beyond 60 hours, changes will eventually be evident in all areas; however, the behavioral significance of the changes will vary. Total sleep loss of 40-48 hours would probably be the upper limit with loss of 30-36 hours more likely. These amounts of sleep loss can be tolerated without debilitating changes in the physiological system. In most instances, if any effects are noted, they will first be evident by changes in mood. Performance changes will be minimal if the tasks are brief, self-paced, highly motivating, and feedback is given as to adequacy of response. Tasks that require sustained vigilance and attention, use of newly acquired skills, retention of new information, and which necessitate long periods to complete are more likely to show sleep-loss effects. Most of the decrement will occur during periods of brief sleep. These effects are more likely to occur during the early morning hours when body temperature is low. Performance workload should be reduced during hours when sleep would normally occur, regardless of actual time of day.

For this paper, the term "sleep loss" will refer to total sleep deprivation, for it is after total sleep loss that physiological, behavioral, and performance changes are most clearly seen. It is also now well established that we can reduce our usual sleep times by 1 to 2 hours for long periods and as much as 3 to 4 hours for 1 to 2 weeks without marked behavioral and performance effects, assuming the usual sleep time was 7 to 8 hours before reduction. Chronic reduction of sleep below 4 hours for 7 to 10 days, however, leads to reduction in effectiveness. Recent studies have shown the recuperative powers of naps and obtaining 6 hours of sleep in 1-hour blocks over a 40-hour period maintained performance in most tasks at the same level as was seen in subjects who obtained 6 hours of uninterrupted sleep at their usual bed time (Mullaney et al. 1983). The once held belief that failure to obtain minimum amounts of a particular type of sleep, Stage REM or Stage 4, would lead to behavioral and mood changes has been abandoned. The amount of sleep one gets is more important than the type. After more than a quarter of a century of sleep research, we still do not know the unique functions of REM and NREM sleep.

More relevant to this paper is the fact that we still do not understand the underlying physiological mechanisms or the function(s) of sleep. So, when we talk about the effects of sleep loss, what is being stressed? Are we measuring the effects of going without sleep, or are the changes due to the increased motor activity and other techniques used to stay awake, i.e., use of stimulants?

Some data are available which suggest that the absence of sleep is the more important stressor. Webb and Agnew (1973) tested the forced rest theory (i.e., rest without sleep would be recuperative), by having one group lie in bed yet remain awake, while another group exercised. After 2 days without sleep, the bed rest group performed as poorly as did the exercised group. Wakefulness was given as the cause of an increase in catecholamine excretion by Steinberg et al. (1968).

Though sleep loss is a significant contributor to resultant changes, one should not assume that the increased activity necessary to remain awake does not contribute to the feelings of fatigue and to performance and biochemical changes. In another bed-rest exercise study, exercise increased performance decrement over that seen in the bed-rest group (Lubin et al. 1976). An interaction between physical work load and sleep loss was found in the level of catecholamine increase as reported by Hasselman et al. (1960).

Physiological Changes

Biochemical Changes. It was initially expected that sleep deprivation, like many other stressors, would result in activation of the pituitary-adrenal cortical axis or in a disruption of its circadian pattern. Conflicting results have been reported. Rubin et al. (1969) reported that although changes were apparent during 205 hours of deprivation in both the amount of activation and in the circadian pattern, these changes were not of sufficient magnitude in each of the 4 subjects to yield statistically significant differences when compared with the predeprivation levels. A definite increase in 17-hydroxycorticosteroids, however, was found during the second half of the 205-hour vigil. Insignificant biochemical changes have been reported by Fiorica et al. (1968) with respect to the catecholamine output. These authors not only found no change in urinary excretion patterns after more than 86 hours of sleep deprivation in 6 young men, but also reported that the ability to regulate body temperature was not impaired by the loss of sleep. It was their conclusion that despite gross psychomotor impairment during sleep deprivation, physiological regulating systems are relatively unaffected by sleep loss.

Froberg and his associates (Froberg et al. 1972), as well as Sternberg et al. (1968) and Hasselman et al. (1960), have reported increased catecholamine output, and Froberg et al. (1972) have indicated the competence of the immune system might be threatened.

Considering all the results to date, it would be safe to say that moderate sleep loss (less than 60 hours), biochemical changes appear to be of questionable reliability and of no major concern for most operational settings.

Electroencephalographic Changes. With sleep loss, there is a decrease in alpha abundance, and the EEG becomes one of low voltage. This low voltage record is similar to that seen in activated alert subjects, as well as in drowsy subjects, and the interpretation of this change has therefore differed. One group maintains that the decrease in alpha is a reflection of increased activation, and a second group interprets the alpha decrement as an indication of a drift toward sleep and decreased activation. The studies of Johnson et al. (1965) and Naitoh et al. (1969) indicate that this electroencephalographic change is a shift toward drowsiness with increasing transient periods of sleep, sometimes called microsleep or lapses, as sleep loss progresses. Neither the Johnson nor Naitoh researchers found an increase in frequencies above 13 cps during sleep loss.

Behavioral Changes

Mood. It is no great breakthrough to state that with increasing sleep loss subjects report increased sleepiness and increased feelings of fatigue, but other feelings also soon become evident. These attitudes influence performance as well as relationships with others. Some adjectives that have been found sensitive to sleep loss are listed in Table 1. A person deprived of sleep usually responds in a negative direction.

Table 1. List of adjectives sensitive to one night of sleep deprivation. (From Lubin et al. 1974.)

Active	Alert
Carefree	Cheerful
Considerate	Efficient
Friendly	Full of pep
Good natured	Lively
Relaxed	Able to concentrate
Able to think clearly	Able to work hard
Dependable	Happy
Kind	Pleasant
Satisfied	

Performance. Performance decrement during prolonged wakefulness has been well-defined, and a consistent picture emerges. Though motivation plays a role, the studies performed at Walter Reed (Williams et al. 1959) demonstrated that the primary impairment during acute sleep loss takes the form of lapses; the subject is unable to maintain efficient behavior and increasingly shows short periods when performance falters or stops. His performance is like a motor that after much use misfires, runs normally for a

while, then falters again. Johnson (1982) summarized performance data and indicated that pacing, length of task, and motivation all determine the extent of performance decrement. In general, when the subject can control the display of stimuli and the rate at which he will respond, his performance will suffer less. Speed will be impaired but accuracy may be maintained. In work-paced tasks, however, so designed because the subject cannot control the time of appearance or duration of the stimulus display and must respond quickly, lapses will cause an increase in errors of omission. Tasks with strong incentive qualities and of brief duration suffer less than tasks that are long and repetitive. The subject's immediate knowledge of results helps to overcome the effects of sleep loss.

References

Fiorica V, Higgins EA, Iampietro PF, Lategola MT, Davis AW. Physiological responses of men during sleep deprivation. Journal of Applied Physiology, 1968, 24,167-176.

Froberg J, Karlsson CG, Levi L, Lindberg L. Circadian variations in performance, psychological ratings, catecholamine excretion, and diuresis during prolonged sleep deprivation. International Journal of Psychobiology, 1972, 2,23-36.

Hasselman M, Schaff G, Metz B. Influences respectives du travail, de la temperature ambiante et de la privation de sommeil sur l'excretion urinaire de catecholamines chez l'homme normal. Comptes Rendus des Seances de la Societe de Biologie, 1960, 154,197-201.

Johnson, LC. Sleep deprivation and performance. Biological Rhythms, Sleep, and Performance, Wilse B. Webb (Ed.), 1982, Chapter 5.

Johnson LC, Slye ES, Dement W. Electroencephalographic and autonomic activity during and after prolonged sleep deprivation. Psychosom Med, 1965, 27,415-423.

Lubin A, Hord DJ, Tracy ML, Johnson LC. Effects of exercise, bedrest and napping on performance decrement during 40 hours. Psychophysiology, 1976, 13,334-339.

Lubin A, Moses J, Johnson LC, Naitoh P. The recuperative effects of REM sleep and stage 4 sleep on human performance after complete sleep loss: Experiment 1. Psychophysiology, 1974, 11,133-146.

Mullaney DI, Kripke DF, Fleck PA, Johnson LC. Sleep loss and nap effects on sustained performance. Psychophysiology, 1983, 20,643-651.

Naitoh P, Kales A, Kollar EF, Smith JC, Jacobson A. Electroencephalographic changes after prolonged sleep loss.

Electroencephalography and Clinical Neurophysiology, 1969, 27,2-11.

Rubin RT, Kollar EJ, Slater GG, Clark BR. Excretion of 17-Hydroxycorticosteroids and vanillylmandellic acid during 205 hours of sleep deprivation in man. Psychosomatic Medicine, 1969, 31, 68-79.

Sternberg H, Guggenheim F, Baer L, Snyder F. Catecholamines and metabolites in various states of arousal. Journal of Psychosomatic Research, 1968, 13,103-677.

Strausbaugh LJ, Roessler R. Ego strength, skin conductance, sleep deprivation, and performance. Perceptual and Motor Skills, 1970, 31,671-677.

Webb WB, Agnew HW. Effects of performance of high and low energy expenditure during sleep deprivation. Perceptual and Motor Skills, 1973, 37,511-514.

Williams HL, Lubin A, Goodnow JJ. Impaired performance with acute sleep loss. Psychological Monographs, 73 (Whole No 14) 1959.

Problems of Using Sleep Research Techniques in the Field

Gerald P. Krueger
Walter Reed Army Institute of Research

Abstract

Experiences in adapting laboratory sleep research methodologies to field study of sustained military operations are reviewed. The use of standard cognitive tests, self-administered questionnaires, an activity monitoring system and observer data collection methods are described.

Introduction

Background. Recent technological advances and modern battlefield tactics give military forces capability to wage war day and night, for weeks at a time, without respite. Such continuous operations require combatants to sustain performance as long as they can with little or no rest and sleep. Such nonstop conditions produce stress, sleep loss, and fatigue, and have important behavioral effects resulting in poor performance, accidents, psychiatric casualties, and reduced mission effectiveness. As a consequence, planners, tacticians, and commanders ask behavioral scientists many questions concerning the impact of sleep loss and the related stresses of sustained operations on performance, and for advice on countermeasures to improve individual and unit effectiveness under such circumstances.

Sleep Deprivation and Sustained Performance. There is a rich literature on laboratory studies of sleep deprivation (e.g. Kleitman, 1939; Webb 1968, 1982); and also on sustained performance during sleep loss (for reviews see Alluisi & Morgan 1982; Krueger, Cardenales-Ortiz & Loveless, 1985; Englund & Krueger, 1985 and Krueger & Englund, 1985). Researchers in military research labs, continue to analyze the nuances of sleep loss and sustained performance variables.

WRAIR's research program on sustained/continuous operations. At the Walter Reed Army Institute of Research (WRAIR), we conduct a research program to predict soldier performance as it is affected by workload, various levels of activity, work/rest schedules and amounts of sleep or sleep loss. Basic behavioral and psychophysiological research is designed to elucidate neuropsychological mechanisms; and field studies serve to identify practical problems and solutions in the conduct of sustained and continuous military operations. Important variables in our program include: prior rest quantity, types of work and tasks, amount of workload, times of day of performance, work/rest cycles, naps, circadian rhythms, diet, physical fitness, individual differences, environmental conditions and the use of pharmacological means to sustain or enhance performance.

Lab Studies. Our lab studies generally focus on cognitive performance and circadian periodicity as in Babkoff et. al.'s, (1985) study of effects of moderate workload and 72 hours of sleep deprivation on a series of perceptual and cognitive tasks. More recently, our lab research program turned to the study of countermeasures to sleep deprivation effects - such as strategic placement of naps, use of hypnotics to induce sleep, and stimulants to sustain or enhance performance even during sleep deprivation. A nap study was conducted in mid-1985, and experiments with hypnotics and stimulants are planned.

Field Studies. There are few data on sustained military performance, or on the frequency and quality of transient sleep periods, in actual training or combat. Questions of who loses sleep, how much, when, why and of what consequence, are important to giving suitable advice to those who ask for it. Our program was therefore expanded to study soldier performance as it is affected by varying work/rest and activity levels, sleep deprivation and sleep discipline in the field, where we can get a more practical view of what really happens in sustained operations. In the past two years we attempted to adapt portions of our lab based methodologies for use in the field by piggy-backing studies onto military test and training exercises: 1) to measure endurance of armor crews while they wore chemical protective clothing; 2) to study command and control personnel in a mechanized infantry REFORGER exercise; and 3) to study military police providing continuous security for heavy weapons storage sites.

Our efforts met with success; but it was clear that much of what is done in the laboratory is not easily adapted to large field studies. All the difficulties of setting up good, well controlled studies, e.g. selection of suit-able performance measurement tasks, reliability and maintenance of instrumentation, complicated scheduling, weather, many soldiers moving about in the macro-environment and scores of uncontrolled variables are amplified in field studies that run 24-hours per day over many days.

Standardized Performance Measurement Tasks for Field Studies. The first problem is selection of suitable performance measures for use in the field. The measures should be sensitive to sleep loss; predict soldier performance as a function of test conditions; not intrude on operational military scenarios; have face validity to soldiers and the leaders who are our customers; and, standardized enough to allow us to compare results among a variety of jobs, task and field conditions.

Off-the-shelf sleep deprivation methodologies were literally inserted into 24 and 72 hour endurance tests of soldiers wearing chemical protective clothing in armor vehicles. A paper and pencil battery of standardized psychological tests (especially cognitive ones), was administered every 3 hours, or 8 times per day. The 20-minute test was given to all pariticpants at battle stations in their vehicles. Our intention was to sample behavior and performance capacity frequently enough to demonstrate circadian or cyclic changes as the operational scenario progressed.

Although most of the armor field tests were scheduled for 72 hours, they frequently lasted only 12-24 hours. In one test, ample rest periods were administratively granted in the actual conduct of the exercise and in another, crew chiefs had soldiers take naps in shifts; thus negating some aspects of sustained operations. Most soldiers did not incur much sleep loss; consequently, it became questionable how valuable our sleep deprivation measures were in identifying performance degradations as a function of sustained operations. Additionally, frequent administration of the 20-min test intruded on the operational scenario.

The military police test had more elements of a true continuous operation. We used trailer mounted microcomputers to administer a 12-min psychological task battery (Thorne et. al., 1985) every 8 hours or 3 times per day. We demonstrated that such computerized data collection can work in the field. However, we still intruded on the operational scenario by taking soldiers away from duty stations to perform our computer tests.

The use of standardized cognitive tasks requires pre-test training so that participants reach some reasonable baseline performance level before the test. It was difficult to obtain sufficient training time (2-3 hrs on each of three days).

Interactions among the amounts of practice, actual rest obtained, novel stimuli, and uncontrolled variables confound the data. However, generally, soldiers performed our tasks well, even while wearing bulky chemical protective suits, and even with moderate amounts of sleep loss.

Our lab based tasks (variously of logical reasoning, addition-subtraction, time estimation, pattern recognition, encoding and decoding etc.) did not readily "fit" into obvious time slots in the military field scenarios. Since the psychological tasks have little face validity, soldiers have difficulty relating to them; and military leaders are not convinced that one can extrapolate findings from these generic tasks to more militarily relevant activities.

Suggestions of using portable "lap computers" to administer such tests may help, but they are not likely to solve the problems identified here. Thus, on the issue of performance measures, we find ourselves back to the argument of whether or not we can/should embed measures into specific soldier tasks and sacrifice the generalizability across jobs, tasks and conditions that the more generic tasks offer.

Self Administered Questionnaires: We used paper and pencil or computerized versions of mood scales (e.g. an affect-activation scale, Genser, 1986), sleepiness scales (e.g. the Stanford sleepiness scale, Hoddes, et. al., 1973) and self-reports of the amount of activity, sleep and rest obtained. They were hardcover booklets for use in tanks, or folded forms for the uniform pocket. Soldiers generally filled out 3-7 minute subjective scales and questionnaires 2-3 times per day.

There were difficulties in getting soldiers to make time to complete questionnaires. In sustained operations, even such a 3-5 minute task, done three times per day, becomes a nuisance.

In one of our studies, soldiers failed to distribute and complete self-report questionnaires in what seemed a passive aggressive response to their dissatisfaction with the field test itself. Pickup and resupply of questionnaires is a challenge when large numbers of soldiers are spread over distance.

Despite accompanying problems, questionnaires are useful to obtain good data in field tests. Reminders include:

1) The research staff, and military unit leadership, must give participants confidence that data will be valuable for a recognizable Army need, and that the participants can contribute to improving the Army.

2) Questionnaires must be simple to complete, quick to administer, and routinely retrieved. Soldiers frequently find too many other things to do and ignore their questionnaires.

3) Give soldiers space on the questionnaire for open-ended comments, and let them know their comments are welcomed. Respond to some comments as the test progresses; get resolution of some of the problems they raise to win their confidence.

Activity Monitoring System. We used an activity monitoring system (Redmond & Hegge, 1985) to gain quantitative information on frequency or quality of transient sleep periods in the field. Soldiers carried in their pocket an activity monitor that recorded body movements on 24-hour tapes. Later, we used wrist activity monitors with 2-week continuous recording capability.

The monitors detect motions of the body, and record counts of movements per unit of time. The wrist monitor is a small accelerometric transducer coupled to a low power microprocessor with a digital memory and is housed in a case the size of a wristwatch. When the data are read and plotted, the resultant actigraph is a time series of activity counts from which one can infer activity, rest and sleep periods.

Such activity monitors are used in sleep labs and clinics; but we have pioneered their use with soldiers in the field. The wear and forget, unobtrusive wrist monitors are accepted by soldiers; and since the data are quite descriptive, the devices offer great promise in accurately measuring activity and sleep patterns in the field. However helpful in giving indications of who is active, who sleeps, and when, the question of "what are the consequences" must be addressed in other ways.

Use of observers: Obviously we cannot fully automate measurement of performance or monitor the activity of large numbers of soldiers. Use of trained observers is still important to describing what goes on in field exercises, and to explaining why. For example, determining that soldiers remain awake doing "busy work" when they could be resting, or determining that personnel in certain units are working harder because they are understaffed are largely judgment calls and only likely to be recorded by observers. Getting good response rates on questionnaires is also enhanced when data collectors relate to soldiers and are participating in the same inconveniences as the troops. Some pointers:

1) Use some uniformed military observers, with rank commensurate to that of the particpants. Observers should spend

time with the participants getting to know their jobs and their alternative ways of doing things.

2) Plan for observers to get rest and sleep too; they tend to miss some unit activity during their own rest period.

3) Post test debriefing interviews of the participants can provide some of the richest data.

References

Alluisi, E.A. & Morgan, B.B. (1982). Temporal factors in human performance and productivity. In: E.A. Alluisi & E.A. Fleishman (Eds.) Human Performance and productivity: Vol. 3 Stress and performance effectiveness. Hillsdale, NJ: Lawrence Erlbaum Inc.

Babkoff, H., Thorne, D.R., Sing, H.C., Genser, S.G. & Hegge, F.W. (1985). Dynamic changes in work/rest cycles in a study of sleep deprivation. Beh. Res. Methods, Instruments & Computers, 17, 5.

Englund, C.E. & Krueger, G.P. (1985). Introduction: Methodological approaches to the study of sustained work/sustained operations. Beh. Res. Methods, Instruments & Computers, 17, 3-5.

Genser, S. G. (1986). Mood scales: Affect and activation in sustained performance studies. (WRAIR Technical Report). Washington DC: Walter Reed Army Institute of Research (In prep.).

Hoddes, F., Zarcone, W., Smith, H., Phillips, R. & Dement, W.C. (1973). Quantifications of sleepiness: A new approach. Psychophysiology, 10, 431-446.

Kleitman, N. (1939). Sleep and wakefullness. Chicago: The University of Chicago Press.

Krueger, G.P., Cardenales-Ortiz, L. & Loveless, C. A. (1985). Human performance in continuous/sustained operations and the demands of extended work/rest schedules: An annotated bibliography. (WRAIR Report No. BB-85-1). Washington DC: Walter Reed Army Institute of Research. (NTIC No. ADA 155-619).

Krueger, G.P. & Englund, C.E. (1985). Introduction: Methodological approaches to the study of sustained work/sustained operations. Beh. Res. Methods, Instruments & Computers, 17.

Redmond, D.P. & Hegge, F.W. (1985). Observations on the design and specifications of a wrist-worn human activity monitoring system. Beh. Res. Methods, Instrumentation & Computers, 17, 5.

Webb, W.B. (Ed.) (1982). Biological rhythms, sleep, and performance. New York: John Wiley & Sons.

Webb, W.B. (1968). Sleep: An experimental approach. New York: The MacMillan Co.

THE ASSESSMENT OF WORKLOAD EFFECTS

Glenn F. Wilson
Armstrong Aerospace Medical Research Laboratory

Abstract

In order to meet the needs of ever increasingly complex Air Force systems, a psychophysiological battery of tests has been developed. The goal of this project is to provide an instrument which can be used to study the effects of stressors in both laboratory and applied settings. The battery is described and sample results are outlined.

As the systems that humans are required to operate become more complex, the level of workload increases. Both mental and physical stresses are increasing in current Air Force systems and no doubt will increase further in future systems. These factors include the large amount of data to be processed, the number of critical decisions that need to be made and the physical effects of high speed maneuvering and long duration missions. Useful information can be obtained during actual missions, from simulated situations and in the design of work stations. Also, physiological measures can be used to provide relatively nonintrusive measures of workload. Several of these measures monitor ongoing activity while others require a secondary task or stimuli which are extraneous to the operator's primary job. By the appropriate selection of tests, it is possible to provide information that is otherwize not available from performance and subjective methods. This information can then be provided to the designer, mission planner or the system that is being controlled by the operator.

At AAMRL we have been interested in situations which manipulate the operator's mental workload. Our efforts have been aimed at finding a number of physiological measures of stress which are sensitive to a wide range of stress types and levels. While no one physiological measure may be appropriate to all situations it is possible to have a battery of tests available which will provide useful physiological measures in many of the situations of interest. Concurrent measures of multiple physiological systems provides not only more information but makes possible the detection of intermeasure patterns and also permits one to simultaneously capitalize on the sensitivities of each measure. A test which is sensitive to fatigue effects can be concurrently recorded with a measure which can show shorter term changes in attention.

A psychophysiological battery has been developed at AAMRL and is called the Neuropsychological Workload Test Battery (NWTB). This computer based system has the capability of presenting stimuli to an operator and recording their physiological and behavioral responses. It can also be used to record nonstimulus locked physiological parameters while the subject is performing other tasks. The battery has an externally triggered mode which permits it to be interfaced with equipment which provides discrete stimuli.

The system uses a PDP 11/73 processor and has a fixed/removable disk, eight channels of analog to digital input, four digital inputs/outputs, a video monitor and headphones for stimulus presentation, a fluorescent tube stimulus display, a graphics terminal and a printer/plotter. This hardware has proven to be reliable in laboratory and simulator environments. On-line data reduction or disk storage modes are available. The raw digitized data and/or the analysed data may be retained. Since a removable format disk is used, long term storage of data is a simple matter.

Eleven tests have so far been included in the NWTB. They are:

1. Memory scanning task
2. Odd-ball, visual and auditory
3. Flash evoked response
4. Tracking rask with evoked response
5. Auditory brain stem response
6. Checkerboard steady state evoked response
7. Sine wave grating steady state evoked response
8. Unpatterned field steady state evoked response
9. Electrocardiogram
10. Electrooculogram
11. Electromyogram

The tests available on the NWTB were chosen from those in the literature that have been shown to have utility for measuring changes due to workload and other stressors. These tests also provide a composite that yields measures of both central and peripheral nervous system acitiity. The development of the NWTB is an ongoing process in which measures are added and deleted as their utility is determined. This permits us to provide the most flexible battery. While tests may be changed, the data format ana the input and output parameters remain the same in order to maintain continuity for the system. The selected measures are tested in our laboratory and in simulators. The unique environment of the simulator provides further constraints upon the tests and the procedures used to implement them. We intend to collect data in actual work stations and use the NWTB to analyse these data.

The NWTB tasks have proven useful in a number of situations ranging from simulator mission segment workload assessment to drug effects. The tests monitor a range of nervous system activities. Basic physiological functioning can be assessed with the auditory brain stem and visual steady state tests while the memory scanning and odd-ball tap higher order cortical functioning. The electrocardiogram (ECG) and electrooculogram (EOG) have been shown to be sensitive to several types of stressors.

Up to three electroencephalographic (EEG) channels can be recorded simultaneously. EOG and muscle activity are used to provide rejection criteria for EEG trials during response averaging. Reaction times and error scores are collected in the appropriate tests.

The software was designed to be user friendly while at the same time permitting flexiblity in terms of parameter selection. Options have been kept to a minimum to simplify the use of the NWTB and to avoid user confusion and errors in set-up.

Several studies have been conducted using the NWTB and have shown it to be a useful device. Data from simulator runs in which the pilot's workload was varied has shown the eyeblink measure to be sensitive to fatigue and attention effects. Blink rate increased as a function of time on task as well as changing with the demands of the mission profile. Heart rate has also been found to be sensitive to the workload level imposed upon the pilot. The visual steady state evoked response is sensitive to fatigue and depressant drug effects. These variables produced changes that suggest slowing of transmission speed with increased levels of fatigue or the drug. The memory scanning and odd-ball tests are also sensitive to drug effects. The visual odd-ball test seems to change with workload demands in simulator environments. The electromyogram as implemented in its first version has not proven to be useful. However, a different method of calculating muscle activity is being programmed and will be tested for its usefulness.

The NWTB development is continuing as we search for new tests and carry on the evaluation of current ones. Further applications of the battery are being tested in an effort to determine the appropriate application of the various tests. We feel that there is need for a battery of psychophysiological tests which can be used by investigators interested in human workload and performance assessment. The goal of our ongoing effort is to provide such a battery and keep it up to date as well as to provide a data base of results. New hardware as well as software will be developed in order to meet the needs of an ever increasing complex work environment.

Subjective Symptoms, Human Endurance, and Cognitive Interventions

T. Michael Rauch, I. Munro, W. Tharion and L.E. Banderet

U.S. Army Research Institute of Environmental Medicine,
Natick, Massachusetts

Abstract

A multivariate model was developed to predict and classify environmentally stressed subjects into survivor and casualty groups. Group prediction and classification functions were based on two separate data bases and therefore two separate linear combinations; one of subjective symptoms and the other of coping strategies designed to mitigate stress. Two highly significant discriminant functions were derived: one from symptom scores and one from coping strategy scores. The symptom classification function correctly classified 94 percent of the cases in their respective group and the coping function correctly classified 98 percent of the cases correctly into the respective casualty or survivor groups.

Introduction

Background:

Individuals engaged in physically and mentally demanding tasks under environmentally stressful conditions (i.e., heat) must often evaluate their energy expenditure, perceived discomfort, and psycho-social contingencies to determine their length and extent of work tolerance. Subjective symptoms or "feeling states" are often related to the endurance of a sustained work period. The point on a continium of work time that an individual expresses perceived discomfort or fatigue may correspond to a similiar point in time which indicates an unwillingness to continue.

The study of subjective symptoms and their relationship as limiting factors in the endurance of work is highly dependent on adequate methods of measurement and techniques to predict and classify individuals into categories of tolerance. The present paper will discuss the results of three studies designed to assess the relationship of subjective symptoms experienced during work in environmentally stressful conditions and their ability to discriminate and predict those willing to endure vs voluntary termination. In addition, a cognitive strategy function which best discriminated casualties from survivors is presented.

Method

Subjects:

The subjects consisted of 81 male volunteers (27 artillery crewmen, 26 armor crewmen, and 28 infantry crewmen) currently on active duty in the US Army. All subjects were medically screened prior to participation in the studies and were required to read and sign a volunteer agreement of informed consent. Subject ages ranged from 18 to 31 years.

Procedure:

Each study assessed the capability of subjects to conduct sustained military field operations (artillery, armor, and infantry) while wearing full chemical protective clothing ensembles in thermally stressful environments (temperatures exceeding 85F). Subjects were administered a battery of psychological tests (including subjective symptoms and a coping strategy inventory) developed by the US Army Research Institute of Environmental Medicine (USARIEM). The test battery was administered prior to test participation (baseline), at scheduled points in time during the field operation, and post field operation. The field exercises were designed to continue for a 72 hour period of sustained operation. Subjects could voluntarily withdraw at any point in time or continue until the exercise was completed or their crew determined combat ineffective by attrition of crewmembers. The 81 subjects were assigned to two post hoc groups ("Casualties" and "Survivors"). Casualties were subjects who voluntarily withdrew from the study during the conduct of field operations; Survivors were required to end their participation due to their crew becoming combat ineffective by crewmember attrition or completion of the operation. Missing data eliminated four subjects from the final analysis, therefore 77 cases were processed (Casualty Group, N=42, Survivor Group, N=35).

Results and Discussion

Two stepwise discriminant analytic techniques were performed on post test data to specify one linear combination of subjective symptoms and one linear combination of coping strategies that maximized the differentiation between Casualties and Survivors. Wilks' lambda was used as the stepwise selection criterion for each analysis. A summary of the stepwise procedure is shown in Table 1 and Table 1A for each discriminant technique. A total of 15 symptoms were selected and produced a high degree of separation as indicated by a final Wilks' lambda of 0.31158, an Eigenvalue of 2.209, a canonical correlation of 0.8297 and associated Chi-Square (15) = 78.71, $p < .001$ for the symptom function. A total of 15 coping strategies produced a very high degree of separation as indicated by the final Wilks' lambda (.19095) and a canonical correlation of 0.899 for the coping function. The Eigenvalue was 4.237 and the associated Chi-Square was (15) = 67.059, $p < .001$. Two classification equations were derived from the pooled within-groups covariance matrix and the centroid for discriminating symptoms and coping strategies respectively. These classification coefficients were multiplied by the raw symptom or coping scores, summed, and added onto the coefficient constant. The equation used for the Survivor group in each analytical technique was:

$$Cs = cs1V1 + cs2V2 + \ldots + cspVp + cs0$$

where Cs is the classification score for the Survivor group, cs1 through cs15 the classification coefficients with cs0 being the constant, and V representing the raw score on the symptom or coping variable. The same equation format was used for the Casualty group with the respective values. Classification scores were converted into probabilities of group membership. The symptom based classification routine was able to correctly identify 93.51 percent of the

cases as members of the groups to which they actually belong and the coping based classification routine correctly identified 98 percent of the cases correctly.

Discriminant scores for a given case were computed by multiplying unstandardized discriminant function coefficients with their respective symptom or coping scores, summed, and added onto the constant. The discriminant scores are of the form:

$$D = d1V1 + d2V2 + d3V3 + \ldots + d0$$

where D is the score on the discriminant function (symptom or coping), d1 is the weighting coefficient for the associated symptom or coping variable, V1 the raw score of symptom or coping variable, and d0 the constant. Symptoms "short of breath", "muscle cramps", "back ache", and "lightheaded" represented the greatest contribution of their associated symptoms on the symptom function. Strategies "I kept going to keep the crew together", "When a situiation kept bothering me, I thought of ways I could control it", "When I was bothered by something I could not control, I did not spend time worrying abcut it", and "I knew there would be a predictable pattern to my alertness and motivation" represented the greatest relative contribution of their associated variables to the coping function. After computing discriminant scores for each case, scores were plotted in a histogram as shown in Figure 1 for the symptom function and Figure 1A for the coping function. The plotting clearly shows the clustering within the two groups, the separation of group centroids and their relative locations, and the small degree of overlap between groups for each function.

As shown above, discriminant analysis yielded an appropriate prediction and classification model. A linear combination of symptoms and one of coping strategies that best distinguished between, and classified into groups, yielded maximum discriminability and probability of group membership with a minimum of measures.

TABLE 1

STEPWISE DISCRIMINANT ANALYSIS SUMMARY FOR SYMPTOMS

STEP	ACTION ENTERED	ACTION REMOVED	VARS IN	WILKS' LAMBDA	SIG.	LABEL
1	ESQ3		1	.65524	.0000	HURTS TO BREATHE
2	ESQ15		2	.59850	.0000	FAINT
3	ESQ26		3	.56772	.0000	NOSE BLEEDS
4	ESQ1		4	.54554	.0000	SHORT OF BREATH
5	ESQ28		5	.53554	.0000	LOST APPETITE
6	ESQ10		6	.52630	.0000	BACK ACHE
7	ESQ12		7	.49848	.0000	LIGHTHEADED
8	ESQ4		8	.47501	.0000	MUSCLE CRAMPS
9	ESQ39		9	.43537	.0000	DEPRESSED
10	ESQ19		10	.41262	.0000	GAS
11	ESQ8		11	.39287	.0000	LEGS OR FEET ACHE
12	ESQ29		12	.37876	.0000	SICK
13	ESQ11		13	.34625	.0000	STOMACHE ACHE
14	ESQ33		14	.32485	.0000	SLEEPY
15	ESQ14		15	.31158	.0000	DIZZY

TABLE 2

CLASSIFICATION FUNCTION COEFFICIENTS FOR SYMPTOMS

(FISHER'S LINEAR DISCRIMINANT FUNCTIONS)

GROUP =	1 CASUALTY	2 SURVIVOR
ESQ1	3.065181	0.1227703
ESQ3	1.048761	-0.2413278
ESQ4	-4.222801	-0.2158528
ESQ8	1.235030	0.3967512E-01
ESQ10	4.367192	0.3105924
ESQ11	-2.075966	0.1441987
ESQ12	-2.277742	0.1713219
ESQ15	1.415299	0.6790291E-01
ESQ29	2.165098	-0.3290753
ESQ33	-0.4469359	0.5970757
ESQ39	1.278873	0.8004203E-01
ESQ14	0.6561753	-0.3155264
ESQ19	-1.101468	0.3335642
ESQ26	-3.749938	1.880635
ESQ28	1.112003	0.1209815
(CONSTANT)	-5.636438	-1.214303

TABLE 1A

STEPWISE DISCRIMINANT ANALYSIS SUMMARY FOR COPING STRATEGIES

STEP	ACTION ENTERED	ACTION REMOVED	VARS IN	WILKS' LAMBDA	SIG.	LABEL
1	COPE6		1	.67798	.0000	KEPT GOING FOR CREW
2	COPE23		2	.57170	.0000	PREDICT ALERTNESS-MOTIVATION
3	COPE13		3	.46725	.0000	CONTROL PROBLEM SITUIATION
4	COPE28		4	.40702	.0000	CDR IN SOCIAL CONVERSATION
5	COPE26		5	.35431	.0000	DIDNT THINK IN ALL OR NOTHING TERMS
6	COPE16		6	.32246	.0000	DIDNT WORRY OVER CONTROL
7	COPE18		7	.29080	.0000	POSITIVE ASPECTS OF SELF-SITUIATION
8	COPE27		8	.25948	.0000	NOT RESPNSBLE FOR WHAT I CANT CNTRL
9	COPE9		9	.24712	.0000	NOT VIEW MISTAKE AS FAILURE
10	COPE7		10	.23607	.0000	DRANK WATER WHEN NEEDED
11	COPE32		11	.22141	.0000	USED IMAGERY/MUSCLE CNTRL TO REV UP
12	COPE20		12	.21200	.0000	DISCUSS PROBLEMS
13	COPE31		13	.20528	.0000	VIEWED MISUNDRSTDING AS MY DIFFCLTY
14	COPE29		14	.19831	.0000	KNEW WHAT TO EXPECT FROM MYSELF
15	COPE11		15	.19095	.0000	DIDNT EXAGGERATE-MINIMIZE PROBLEMS

TABLE 2A

CLASSIFICATION FUNCTION COEFFICIENTS FOR COPING STRATEGIES

(FISHER'S LINEAR DISCRIMINANT FUNCTIONS)

GROUP =	1 CASUALTY	2 SURVIVOR
COPE6	3.385104	9.457964
COPE7	9.652900	11.38394
COPE9	-2.850169	-4.266487
COPE11	-0.6139346	0.6493284
COPE13	1.821716	-2.802655
COPE16	3.465747	6.508793
COPE18	6.455670	9.133534
COPE20	-1.776882	0.8079285
COPE23	5.798568	8.772315
COPE26	0.5857569E-01	-1.971056
COPE27	1.566764	-0.1469537
COPE28	2.958473	5.408597
COPE29	3.664583	2.287862
COPE31	-1.908375	0.1063051
COPE32	-1.902878	-4.673403
(CONSTANT)	-34.73853	-57.39615

FIGURE 1

ALL-GROUPS STACKED HISTOGRAM
CANONICAL DISCRIMINANT FUNCTION SYMPTOM

```
       16 +
          |
          |                                2
 F        |                                2
 R     12 +                                2
 E        |                                2
 Q        |                                2
 U        |                                2
 E      8 +                                2
 N        |                                2
 C        |                                2
 Y        |                                2  2
        4 +                                2  2           1
          |                            2   2 22      2 11111
          |                           22 2 2221   2 1111111111111 1
          |                       2  222 2212212 211111111111111111  11
           X---------+---------+---------+---------+---------+---------+--------
         OUT      -6.0      -4.0      -2.0        .0       2.0       4.0
22222222222222222222222222222222222222222111111111111111111111111111111111111111
CENTROIDS                                 2              1
```

FIGURE 1A

ALL-GROUPS STACKED HISTOGRAM
CANONICAL DISCRIMINANT FUNCTION COPING

```
        8 +
          |
          |
 F        |
 R      6 +                          1
 E        |                          1
 Q        |                          .
 U        |                          1
 E      4 +                          1 1
 N        |                          1 1
 C        |                          1 1               2   22  2 2
 Y        |                          1 1               2   22  2 2
        2 +                         11 11       1  2   2 2 22  2222
          |                         11 11       1  2   2 2 22  2222
          |                       1111 111  111 11 12  222 22  2222  2 2
          |                       1111 111  111 11 12  222 22  2222  2 2
           X---------+---------+---------+---------+---------+---------+--------
OUT      -6.0      -4.0      -2.0        .0       2.0       4.0
11111111111111111111111111111111111111111222222222222222222222222222222222222222
 CENTROIDS                               1                   2
```

SYMBOLS USED IN PLOT

SYMBOL	GROUP	LABEL
1	1	CASUALTY
2	2	SURVIVOR

PROBLEMS IN MODELING COMBAT STRESS: A PROGRAM TO MEET THE CHALLENGE

Gerald A. Hudgens[1], James P. Torre, Jr.[1], Robert T. Chatterton[2], Samuel Wansack[1], Linda T. Fatkin[1], Frank DeLeon-Jones[3]

Abstract

Since most military equipment is created for human use under combat conditions, there is an important human factors requirement for basic human and weapon-systems performance data obtained under "combat-stress" conditions. Data from actual combat conditions are virtually nonexistent, and the data which do exist have not, for the most part, been collected under combat-like conditions. Many problems hinder attempts to obtain performance data under conditions mimicking combat stress: e.g. defining the elements and nature of combat stress, mimicking stressful aspects of combat without endangering test participants, measuring physiological and psychological stress effects under test conditions, overcoming test sophistication, creating realistic simulations, and dealing with physiological and psychological factors which result in individual patterns of response to stress. A multi-phase, multi-disciplinary research program has been initiated to deal with these problems and to lead the development of standard operating procedures for collecting performance data under conditions that can be reliably identified as highly stressful to the test participants.

Introduction

It is the function of the human factors discipline to be concerned with the performance capabilities of humans as operators of equipment, to be concerned with those aspects of equipment design which will allow human operators to use the equipment, and to be concerned with the interaction of these factors to maximize the performance of the human-equipment system when it is used for its intended purpose. The Human Engineering Laboratory and numerous other R&D agencies spend much time, money and effort each year collecting data relative to design and performance problems. These data are usually collected in relatively benign laboratory experiments, field studies, "demonstrations," and range tests, most of which are primarily equipment tests which seldom attend to situational conditions that might affect the operators' abilities to perform at or near their potential. The extent of the problem of ignoring these factors is most evident when we realize that much or most of the equipment that we in the DoD are concerned with is intended for use under conditions of "combat stress."

[1] US Army Human Engineering Laboratory, Aberdeen Proving Ground, MD
[2] Northwestern Memorial Hospital, Chicago, IL
[3] VA West Side Medical Center, Chicago, IL

It would seem logical that during the R&D process performance data should be collected relative to the most extreme conditions in which the system might be expected to operate -- under combat-like conditions. Furthermore, the requirement for data collected on human-equipment systems exposed to combat-like conditions is bolstered by the needs of systems analysts and operations researchers to estimate systems performance in models of combat situations. The lack of such data for making accurate estimates in the models has frightening implications regarding the decisions that could be based on the models' predictions.

It is the intent of this paper to discuss many of the problems standing in the way of collecting data under combat-stress conditions, and to outline a program designed to provide a means for collecting data under stress conditions approximating those of combat.

Problems

The Nature of "Combat-Stress." Certainly, one of the biggest problems has to be that of defining the nature of combat stress and determining all of its elements. Probably the one element of combat which would be considered most universally and severely stressful is the threat to life and limb. Aside from waiting for the next war, however, this is the one element we cannot incorporate in combat-stress scenarios for collecting performance data. Other probable aspects of the combat-stress situation, which could be more easily controlled or manipulated in test scenarios, include: peer pressure to perform for the welfare of the unit, the degree of preparedness to meet the demands of the situation, the amount of information about the specific situation provided in advance, physical conditioning, the degree of physical or mental effort required in the situation, and amount of rest provided or withheld. It is through the use of some combination of these factors, within the current human-use guidelines, that we hope to be able to establish a set of standard operating procedures (SOPs) which reliably produce severe stress responses in test participants (TPs) and which can be readily used in laboratory and field test situations.

Test Sophistication and Lack of Realistic Simulation. Simulations, by definition, are not real; however, technological advances allow us to continue improving our simulations. It is ironic, however, that, at the same time, the widespread use of these same technologies is making our subject pool more sophisticated about them, and lessening their effectiveness. Additionally, the increased testing, and publicity about it, in our society has made people quite cognizant of experimenter limitations and subjects' rights to remove themselves from test situations for any reason without penalty, both factors which tend to change the stress of the test situation.

Measuring Stress Response. Responses to stress take a variety of forms, both psychological and physiological, and methods have been developed for measuring both kinds. Psychological measures depend either on clinical evaluations, which are subject to individual and theoretical bias, or on self reports which are limited by individuals' abilities of expression, memory, honesty, etc. Both the validity and reliability of psychological assessments of response to stress are often questioned. More

generally accepted have been a variety of physiological meaures including: catecholamine, hormone, and peripheral measures such as heart rate and galvanic skin response. However, when multiple physiological measures have been obtained on individuals exposed to stress, ambiguous or contradictory indications regarding stress responsiveness are often obtained. Further confounding the problem of physiological indicies, is the fact that so many measures have proven to be sensitive to stress that one cannot hope to employ all of them in a particular study, and little information exists as to which measures or combination of measures might be best in any given situation.

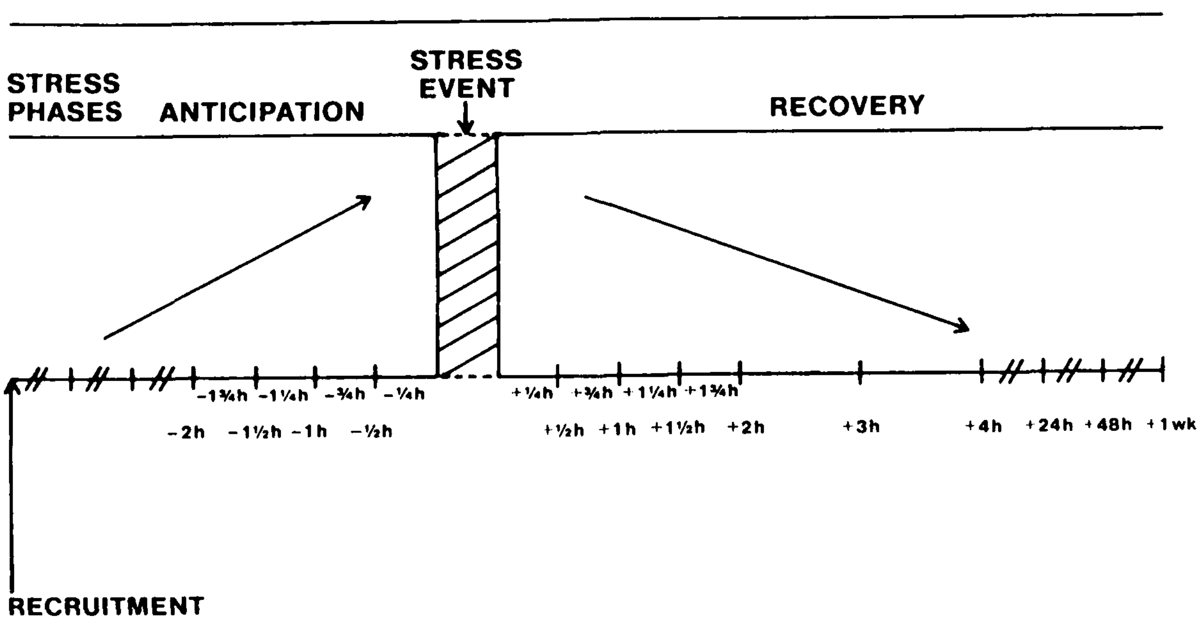
Individual Variation in Stress-Response Patterns. Variation between and within individuals in response to a given stressor and across different stressors has led to inconsistent findings and much confusion in the literature, largely because such findings were at variance with Selye's notion of a generalized-adaptation-syndrome which has so influenced the stress area for the past half century (Selye, 1936). More recent research and theorizing has begun to account for much of the variability and to raise the prospect of a new approach to the study of stress based on factors underlying this variability. Individuals' perceptions of stress appear to be quite important. Mason (1975) demonstrated that the perception of stress is critical to stress response: a physical stressor alone is insufficient. And the importance of personality traits and coping styles has been emphasized by Rose (1980). The importance of various aspects of the physiological indicies, themselves, has also been indicated. Mason (1974) and Fibiger & Singer (1984) have demonstrated that different kinds of stress result in different response profiles when more than one index is used. And Swenson & Vogel (1983) have shown that the severity of the stressor can influence the response profile by affecting the duration of adrenal-cortical response. It is the prospect that stress-response profiles, obtained under different kinds and levels of stress conditions, might yield a crude reference scale against which the profiles obtained under our test conditions might be compared, that has encouraged us to pursue the following program.

Program

Rationale. A program has been initiated to develop a set of SOPs for reliably evoking severe stress responses in TPs while performance data are collected in both laboratory and field test settings. The question of relating these data to performance in actual combat stress will necessarily have to be postponed until the opportunity exists to validate with comparable data collected under combat conditions. In the mean time estimates of performance under combat stress will be significantly upgraded over current estimates based on non-stress test conditions. To be successful we will need, first, to have the capability of knowing when and to what extent we have stressed a person. With this capability, we then can begin to manipulate various psychosocial variables, within human-use guidelines, to induce stress until we have established a set of SOPs which reliably produce stress profiles with elements typical of those common to other kinds of severe stress such as those studied in this program. For the present, the program will be limited to investigating acute stress conditions.

Phase I. In this phase, which is currently underway, it is our intent to obtain extensive stress-response profiles for at least two levels of stress in four protocols representing four distinct categories of stress: 1) Examination stress -- male medical students taking a written class quiz of moderate importance to a class grade or taking oral exams of critical program importance before rating peers and professors; 2) Surgical stress -- male spouses of patients facing and undergoing either a diagnostic laparoscopy procedure or a surgical operative procedure for cancer; 3) Film stress -- male and female college students viewing a benign film, a film of staged woodshop accidents, or a film depicting actual aboriginal ceremonial rites of crude operations to the penis and scrotum of adolescents; 4) Salvo stress -- a field study, in which male Army and Marine riflemen will fire for record on a rifle range either under standard conditions or under highly competitive conditions in which we will pilot test several psychosocial variables as stress inducers. The figure below represents a generalized time frame for the four protocols. The following procedures generally apply to all protocols, with some modifications as practical considerations dictate.

TIME CHART FOR STRESS PROTOCOLS



At about the time of recruitment, the TPs are given a battery of standardized personality tests to assess degree of extroversion, sensation-seeking, locus of control, and generalized anxiety. At about recruitment and/or at about one week or more after the stress event basal blood samples are obtained. Additional blood samples are generally obtained at 15-min intervals from -2h to +2h around the stress event and at other times as dictated by the anticipated recovery period. The samples will be assayed for epinephrine, norepinephrine, cortisol, testosterone, prolactin, growth hormone, TSH, free triiodothyronine (T3) and luteinizing hormone. A battery of standardized anxiety/stress-perception and coping

measures are given shortly before and immediately following the stress event.

Phase II. This phase, which will begin with the pilot field test in the spring of 1986, is the phase in which we will test, in laboratory and field studies, the effectiveness of various motivational and psychosocial variables to increase significantly the level of stress reported by TPs and reflected in their physiological stress profiles. These profiles will be compared with those obtained in Phase I for the purpose of validation of the existence and degree of stress elicited.

Phase III. This phase will conclude the SOP-development aspect of our program for acute stress conditions by demonstrating their effectiveness in a performance evaluation of a major weapon system.

Phase IV. This phase represents further directions for the program: 1) the evaluation of various intervention techniques thought to improve performance (e.g., physical conditioning); 2) the study of factors leading to the complete suppression of appropriate performance and of ways of coping with suppression; and 3) the expansion of our efforts into investigations of the effects of multiple stressors and chronic stress on stress-response profiles and on performance.

The program described above is clearly an ambitious one. It has as its immediate goal the solving of an important Army problem -- how to obtain performance data with relevance to combat application. It has, moreover, the potential of providing data which could contribute significantly to the issues of measuring stress responses and individual differences in stress responsiveness.

References

Fibiger, W. & Singer, G. (1984). Physiological changes during physical and psychological stress. Australian Journal of Psychology, 36, 317-326.

Mason, J. W. (1974). Specificity in the organization of neuroendocrine response profiles. In P. Seeman and G. Brown (Eds.) Frontiers in neurology and neuroscience research. Toronto: University of Toronto.

Mason, J. W. (1975). A historical view of the stress field: Part II. Journal of Human Stress, 1, 22-36.

Rose, R. M. (1980). Endocrine responses to stressful psychological events. Advances in Psychoneuroendocrinology, 3, 251-276.

Selye, H. (1936). A syndrome produced by diverse nocuous agents. Nature, 138, 32.

Inhibition of Motion Sensitive Visual Units by Stationary Patterns

by
William P. Marshak and Rex O. Gibson
United States Air Force Academy

Abstract

Several models of motion perception (Marshak and Sekuler, 1979; Marshak, 1981; Marshak, 1986) propose that visual motion creates excitation over a large numbers of directionally tuned units in the visual system. An inhibitory mechanism is proposed to narrow this excitation so that direction of motion can be accurately determined. The present study provides evidence of such inhibition by showing motion aftereffect is attenuated by stationary patterns in proximity of adapting motion. These results support the idea that background stationary patterns inhibit the activity of motion sensitive units.

Introduction

The visual system is often portrayed as being comprised of independent channels dealing with form, motion and color. Yet the same neural pathways carry these various types of information and it is unreasonable to believe no interaction occurs. Interactions between these various channels have already been reported.

Riggs and Day (1980) observed a dramatic interaction of color and motion when they adapted subjects to two different colored patterns moving in two different directions. Color contingent motion aftereffects were observed. When a test pattern identical to one of the adapting stimuli was presented, the appropriate aftereffect was observed. The same was true for the other adapting stimulus. Presenting achromatic test patterns produced a summing of the aftereffects. A second interaction, this time of motion and pattern was reported by Sekuler and Levinson (1978). When they spun disks covered with radial patterns with missing gaps, subjects perceive closure of the gaps and solid radial patterns. The motion of the stimulus creates illusory patterns.

Marshak and Sekuler (1979), Marshak (1981) and Marshak (1986) have proposed that moving stimuli generate broadly tuned excitation in the visual system. The latest version describes how energy in the temporal domain is generated over large angles by objects in motion. The temporal sensitive elements of the visual system then must interpret this energy into direction and velocity. This excitation must be modified by a proposed inhibitory

mechanism before it can serve as a basis for determining the direction of motion. One established source of inhibition is a spatially overlapping pattern moving in a different direction. Marshak and Sekuler (1979) reported that patterns moving in different adjacent directions were perceived to repulse each other. Their perceived directions were deflected away from the other direction of motion.

Overlapping directions in the same portion of the visual space are very unusual stimuli. Under most conditions, only one direction of motion is perceived in one portion of the visual space at one time. What then is the source of inhibition under more typical viewing conditions? One possible candidate might be inhibition generated by spatial information in the adjacent background. Could the spatial channel be a source of the inhibition?

One paradigm for studying the operation of the motion sensitive channel is the motion aftereffect (MAE) illusion (Adams, 1831). When moving patterns drift through the visual field then becoming stationary, an observer perceives motion in the direction opposite of adaptation. Sutherland (1961) attributed the aftereffect to "fatigue" of neurons sensitive to the adaptation direction of motion. These neurons enter a refractory period of reduced activity following adaptation, so that unadapted neurons (sensitive to the opposite direction) seem relatively more active and motion opposite to adaptation is perceived.

If depleting the firing capacity of neural elements is the basis for MAE, then generating inhibition during the adaptation period ought to prevent neural "fatigue" in the cells affected by the motion and attenuate the aftereffect illusion. Observing such an attenuation of MAE provide indirect evidence that real motion perception can be affected by stationary patterns. Marshak (1984) demonstrated that inhibition could occur with overlapping dot patterns. The stimulus was analagous to the overlapping moving patterns of Marshak and Sekuler (1979). The current experiment determines if adjacent stationary patterns have a similar influence.

Methods

Subjects were all U.S. Air Force Academy cadets between the ages of 18 and 24 years. There where 46 men and 4 women who had 6/6 vision or wore correction to that value. Subjects were all volunteers with no experience as psychophysical observers. Each subject was randomly assigned to the five conditions of the experiment.

Five different stimulus disks were used to produce the MAE. Each disk was white with a diameter of 21.2 cm. Viewing distance was 57 cm so that the disk subtended 21.2 degrees of visual angle. Each disk was connected to a Bodine Electric Company motor with a silver center nut of 3 cm diameter. Rotation during adaptation was at a rate of 13 rotations per minute or 78 degrees per second.

The basic configuration of the stimulus disks were 12 radii lines at 30 degree spacing each subtending .5 degrees of visual angle. In the four experimental conditions there were 0,1,2 or 3 circles with the same width as the radii. When present, the one circle was located at 12 degrees of visual angle from center. Two circles were located at 6 and 12 degrees from center. Three circles were located at 6,12 and 18 from center. The rings, although rotating with the radii, were stationary contours in the field of motion. There was one additional control condition to determine if reduced adaptation area of the stimulus influenced the afterefect. In this condition, there were interruptions in the radii of width .5 degrees of visual angle at 6,12 and 18 degrees from the center of the disk.

Subjects viewed the stimuli at eye level with illumination provided by overhead florescent fixtures. Illumination from overhead florescent lights was measured at 7.2 cd/meter squared and contrast between black lines and white background was approximately 80 percent. Adaptation was always for 60 seconds. Instructions were to fixate on the center of the disk through adaptation and upon stopping the stimulus, watch the illusory motion and report when all motion stopped. Data from four subjects were discarded and replaced with four randomly assigned replacements. Two subjects were dropped because of equipment failure, one due to an experimenter error and one due to a failure to follow instructions.

Results

The means and their standard errors of the four experimental and one control conditions are plotted in figure 1. As you can see, their is reduced MAE with increased presence of stationary circles. The four experimental conditions were analyzed by one way analysis of variance with trend analysis. The analysis revealed a linear trend (Fcalc= 7.087, Fcrit{1,30 df,alpha= .05}= 4.07). Presence of one circle reduced MAE by 24.8%, two rings resulted in a 30.2% decrease and three rings a 32.1% decrease.

The reduction of MAE has an alternative explanation. Perhaps the presence of stationary circles reduced the retinal area over which adaptation took place. To test this alternative, post-hoc Newman-Keuls tests were performed between the control condition with the three concentric disruptions of the adapting radii with the 0 and 3 ring position. The control condition was not significantly different from either of the experimental conditions (both tests had Qcalc = 1.88, Qcrit {r=1, df within= 30}= 2.89 and Qcrit {r=3, df within= 30}= 3.49). The disrupted radii control condition produced a 16.1% reduction in MAE, only half as much as the concentric line stimuli. The overlapping stationary patterns of Marshak(1984) did not reduce the area of the adapting stimulus, so we can conclude at least half of the MAE reduction came from a source other than reduced area.

Discussion

The increasing presence of stationary concentric lines results in decreasing amounts of motion aftereffect. This effect cannot be accounted for solely by reduction in adapting area. It is reasonable to conclude that somehow motion sensitive units in the visual system are inhibited during adaptation by the presence of stationary patterns and do not suffer as much adaptation because of the inhibition.

These finding extend those of Marshak (1984) in which the presence of stationary dot patterns overlapping moving dot patterns also decreased MAE from adapting stimuli without stationary dots. The current results show the same effect can occur without spatial overlap of the stimuli.

The inhibition of MAE by stationary patterns implicates adjacent pattern information as an influence on motion perception. It is plausible that spatial information in the surround may limit the range of angles that temporal excitation operates and provides an accurate basis for the perception of motion's direction.

Bibliography

Adams, R. (1894). An account of a peculiar optical phenomenon seen after looking at a moving body, etc. Lon. and Edib. Philos. Mag. and Sci. (3rd series) 5, 374-377.

Marshak, W.P. and Sekuler, R. (1979). Mutual repulsion between moving visual targets, Science, 205, 1399-1401.

Marshak, W.P. (1981). Perceptual integration and differentiation of directions in moving patterns, Doctoral Dissertation, Northwestern University.

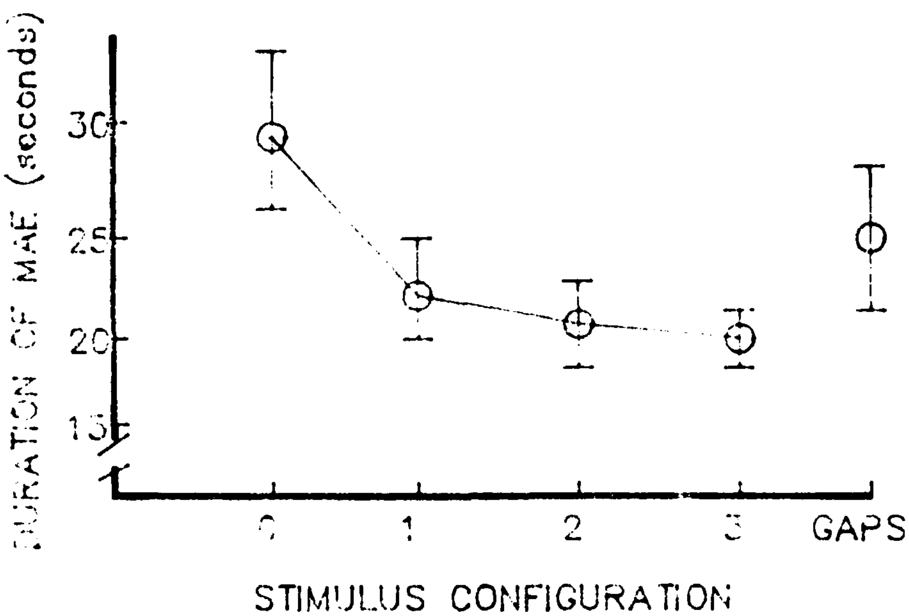
Marshak, W.P. (1984). Interactions of motion perception with pattern perception, Proceedings of Psychology in the Department of Defense, AF-TR-84-2, 682-685.

Marshak, W.P. (1986). Unpublished research report.

Riggs, L.A. and Day, R.H. (1980). Visual aftereffects derived from inspection of orthogonally moving patterns, Science, 208,416-418.

Sekuler, R. and Levinson, E. (1978). The perception of moving targets, Scientific American, 236(1), 60-73.

FIGURE 1: MAE AS A FUNCTION OF STATIONARY CONTOURS OR GAPS.



EFFECTS OF WIDE CAMERA BASELINES AND MAGNIFICATION ON VISUAL PERFORMANCE WITH A STEREOSCOPIC TV SYSTEM

Edward H. Spain and Ross L. Pepper
Naval Ocean Systems Center - Hawaii Laboratory

Abstract

A dual-channel video recording and playback system was constructed in order to investigate the effects of several key sensor parameters on visual performance with a stereoscopic TV display system. It consisted of a synchronized pair of optical video disk recorders under computer control which was used to record stereoscopic still video frame-pairs in the field and play them back in a controlled laboratory environment for visual performance data collection. An experiment was conducted to assess the independent effects of camera interaxial separation, magnification, and their simultaneous interaction on target detection and recognition. The results of this experiment suggested that both image magnification and increases in camera interaxial separation are useful strategies for enhancing target detection time and recognition rate. The interaction of these two factors did not disrupt performance.

Introduction

The Naval Ocean System Center's Hawaii Laboratory is committed to research supporting both undersea and terrestrial applications of teleoperators and other man-in-the-loop systems. Earlier research efforts (Pepper, Smith, and Cole, 1981; Pepper, Cole, and Spain, 1983) established the usefulness of stereoscopic (stereo) TV displays for teleoperators. Follow-on studies (i.e., Spain and Cole, 1982; Spain, Cole, and Pepper, 1984) using abstract test stimuli identified camera interaxial separation, magnification, and motion parallax as three key factors affecting visual performance with such systems. The effort reported in this paper extends past research by determining effects of two of the three key factors on target detection times and recognition rates for targets set in cluttered, natural scenes.

Stereo TV systems are simple in basic conception and easily constructed out of widely available hardware, but they pose serious difficulties for operators in many real-world applications. Stereoscopic perception under direct viewing conditions (i.e., those in which observers are physically present in the scene of interest) has been studied scientifically for well over 100 years and remains an active area of investigation among vision researchers with many central questions remaining unresolved. Viewing a remote scene through a stereo TV system has been investigated scientifically only in the past two decades and remains less well understood. The scenery conveyed by a stereo TV system may be separated from an observer by any distance to which remote video cameras and lines of communication can be extended. What is seen through such systems is frequently unfamiliar and poses substantial perceptual and cognitive challenges for an operator.

The objective of remote terrestrial reconnaissance is to see adversarial forces without being seen or at least from a position out of range of hostile fire. This involves viewing across considerable distances so that the images of significant objects subtend very small visual angles, frequently below the threshold for acuity with unaided vision. Optical gunsites, telescopes, field glasses, and binoculars are all standard military hardware because they effectively extend the range and targeting accuracy of unaided human vision. Magnification simply enlarges the images of distant objects to retinal angles for which the human visual system is adapted to extract meaningful information. Though the user of such devices may not be consciously aware of it, magnification frequently produces large distortions of textural, density, relative size, and linear perspective cues. In general, the greater the magnification, the greater the distortion of this important class of visual cues to depth, distance, orientation, and direction. How such distortion influences perceptual judgments has primarily been studied under monocular or monoscopic viewing conditions in the laboratory with abstract stimulus patterns. The distortion reportedly becomes much more apparent and distracting to some observers when magnified scenes are viewed stereoscopically. This effect has generally been attributed to the somewhat nebulous idea of conflict between two classes of visual cues - retinal disparities and textural gradients.

Given optimal viewing conditions, the normal range for stereopsis extends only out to about 450 meters for unaided viewing, far short of the distances at which militarily significant objects can be seen with unaided vision. Extending the effective range of stereopsis through a binocular viewing system is straightforward. One simply extends the lateral separation between an observer's virtual viewpoints. In addition to providing a 7X to 10X image magnification, most field binoculars widen the virtual viewpoints of the observer's eyes to about twice the normal interocular separation. This doubles the disparities of objects within the binoculars' field of view, and it doubles the range of effective stereopsis. Larger separations are constrained by practical considerations of size, weight, and image stability. Present-day binoculars represent a host of practical compromises which have evolved over centuries of use. Little in the way of controlled visual performance testing (as opposed to routine optical testing) has been conducted to investigate optimal configurations of various interaxial separations for various perceptual tasks.

Unburdened by the physical constraints imposed by binoculars, a pair of stereo TV cameras is readily separable by many multiples of the normal binocular interaxial separation, expanding the range of the stereoscopic field to many times the normal range of stereopsis associated with binoculars. Consequences of widening interaxial separation on the retinal pattern of stimulation are predictable on the basis of geometrical models of stereoscopic transmission through imaging systems. Likewise, the effects of image magnification on proximal stimulation are predictable by means of straightforward geometrical models. What remains largely unresolved and is vitally important to the development of remote stereoscopic viewing systems are the independent perceptual consequences of the image transformations produced by magnification and camera interaxial separation especially when other artifacts are present in images due to the limited image fidelity of available video systems.

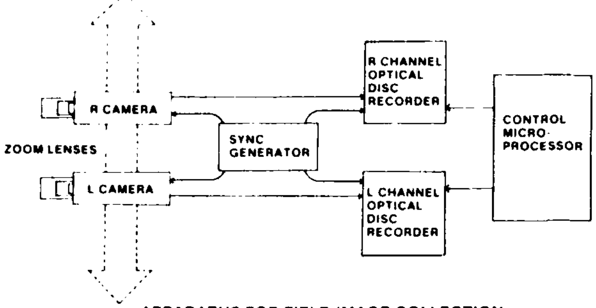
Even less well understood, but probably even more important, are the perceptual consequences of simultaneously varying both interaxial separation and image magnification for a remote reconnaissance task using three-dimensional targets set in cluttered natural scenes. The research reported here provides measures of visual performance (i.e., target detection time and target recognition rate) under a variety of camera interaxial separations and magnification values. It directly addresses the question of a possible interaction of camera interaxial separation and magnification on both target detection times and recognition rates.

Methods

Procedures and apparatus used during the course of experimentation can be divided into two categories: 1) those used for stereo image collection, and 2) those used for visual performance testing. Video recording was required to ensure comparability of the complex, outdoor imagery used in all presentations to all experimental observers. Repeated playback of the same video frames over the course of testing sessions and to the various observers required that the recorded image information not be degraded by repeated access for playback. Both instrumentation requirements were met by a pair of synchronized Panasonic TQ-2023 optical memory disk recorders (OMDR's). The OMDR's provided real-time video recording capability at resolution higher than that attainable with portable video tape recorders. Figure 1 presents a schematic diagram of the video system used for stereo image collection.

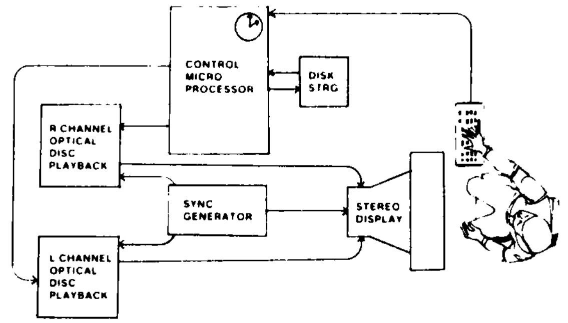
Since outdoor scene conditions constantly change due to factors beyond experimental control, it was necessary to take a series of stereo picture pairs from varying interaxial separations in rapid succession before conditions in the natural scene changed substantially. This provided essentially identical stereo images for all effects other than interaxial separation. An array of four unequally-separated cameras, each with its own zoom lens and pan, tilt, and roll base was constructed and level-mounted on a flatbed truck. Optical axes of all four cameras were paralleled. The camera array was elevated 2 meters (m) above ground level during stereo image collection. Video signal output levels were electronically matched for all four cameras prior to image collection. Unequal spacing of the cameras in increasing multiples of the average human interpupillary distance (I=63.5mm) provided five separate left-right camera pair interaxial combinations. By time-multiplexing the video signal outputs from these camera pair combinations to the left and right channel OMDR's, five separate stereo views of a scene were taken in rapid succession.

Figure 1
VIDEO HYPERSTEREO VIEWING



APPARATUS FOR FIELD IMAGE COLLECTION

Figure 2
VIDEO HYPERSTEREO VIEWING



VISUAL PERFORMANCE TESTING APPARATUS

Two types of camouflaged targets, either a lone, standing man or a jeep were presented against a cluttered, outdoor background. Equal numbers of video frames were displayed in which the man or the jeep or no targets were present. In addition, order of presentation for the man, jeep, and no target trial types was randomized so that there was a .333 probability of presentation for each target type on any given trial. Prior to testing, observers were informed that the order of presentation for trial types would be randomized, and therefore unpredictable. Target position within the display field of view was counterbalanced across three positions (i.e., left, center, and right) in a testing session so that observers would not develop positional searching biases. For each target type, equal numbers of trials were presented in which linear distance from the cameras to the target was either 200m, 400m, or 600m. A total of 120 different video frames was collected and displayed during testing.

Seven young adult observers (three females and four males) participated in the experiment. All observers were within normal ranges for Snellen acuities, phorias, and measures of stereopsis. Figure 2 presents a diagram of the visual performance testing apparatus. Both two-dimensional and three-dimensional images were presented to observers by means of a bench-mounted polarizer stereo TV display (see Cole, Pepper, and Pinz, 1981 for a detailed description and illustration). Display intensity and contrast were balanced subjectively prior to each day's testing with average luminance available at the observer's eye equal to approximately 10 foot-Lamberts. An observer was seated in front of the display in a darkened room with the index and middle fingers of the right hand resting on a pair of high-speed telegraph keys and the right foot resting on a foot pedal. Head position and polarizing filter orientation were constrained by use of a viewing hood. Eye to display screen distance was fixed at .5 meter, establishing an overall display visual angle of 40 degrees by 30 degrees.

At the beginning of each trial, a video frame was displayed in which the word "READY" was centered on the screen at screen depth (i.e., with zero disparity). This was the observer's cue to start the trial sequence whenever she/he was prepared to do so by tapping the foot pedal. After a randomly varying delay of from .5 to 1.5 seconds following the foot tap, the response time clock started simultaneously with the onset of a pre-selected video frame. The observer was instructed to tap either or both of the telegraph keys upon detecting a target in the video image. Elapsed time between stimulus onset and the observer's key tap response was recorded. Immediately following the key tap, the words "MAN" and "JEEP" were displayed at screen depth. This prompted the observer to indicate which target was present in the scene by tapping the left (index finger) telegraph key to indicate the man and the right (middle finger) key to indicate the jeep. This response was not timed. For trials in which no target was presented, the observer was instructed to refrain from tapping the telegraph keys. On these trials, four seconds elapsed before scene display was terminated. Immediately following the observer's response, a frame indicating whether the response was either "CORRECT" or

"WRONG" was displayed for one second.

Interaxial separations of twice the average interpupillary distance, i.e., 2I(130mm) as well as 4I(260mm), 6I(390mm), 10I(1300mm), and 12I(1560mm) were completely crossed with magnifications of 1X(40 degrees), 1.25X(32 degrees), 2X(20 degrees), and 4X(10 degrees) in the experimental design. Target distance and position in the display FOV were counterbalanced for the various combinations of camera interaxial separation and magnification. Each observer participated in 3 sessions of approximately 90 minutes duration each. Sessions were held on separate days and each session consisted of 2 blocks of 240 trials for a total of 480 trials. Each observer was tested on a total of 1440 trials.

Results

Target detection times and target recognition rates were subjected to a repeated-measures analysis of variance. A statistical significance level of $p<.05$ was set a priori for all effects in all analyses. Conservative statistical procedures were used to hold experimentwise Type I errors to less than the stated significance level. All analyses revealed large individual differences between observers. The main effect of camera interaxial separation on target detection times is plotted in Figure 3. This effect was statistically significant ($F(4,20)= 3.41$, $p< .05$). A significant main effect for magnification ($F(3,15)= 20.0$, $p< .01$) on target detection time was also found. No significant interaction between camera interaxial separation and magnification was found for target detection time. A significant main effect for camera separation on target recognition rate ($F=(4,24)= 19.39$, $p<.01$) is plotted in Figure 4. A significant main effect of magnification ($F(3,18)=71.12$, $p<.01$) on target recognition rate was also found. More importantly, a significant interaction ($F(12,72)=4.98$, $p<.05$) between camera interaxial separation and magnification was found for target recognition rate.

Figure 3

EFFECT OF CAMERA SEPARATION ON DETECTION TIME

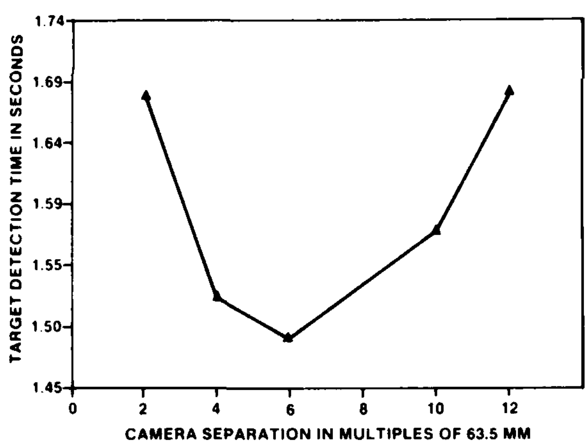
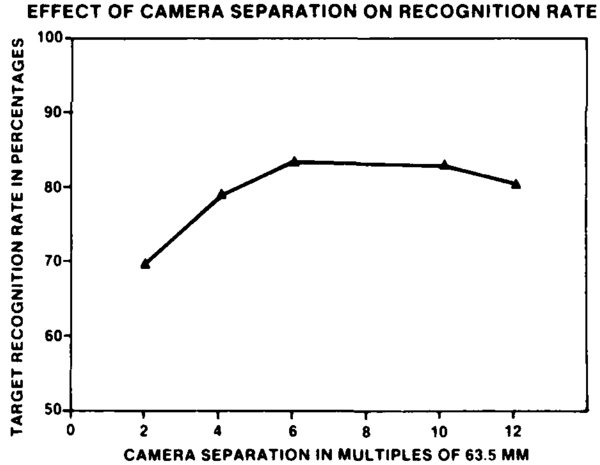


Figure 4

EFFECT OF CAMERA SEPARATION ON RECOGNITION RATE



Discussion

The effects of image magnification on performance correspond well with the available literature on target acquisition with TV displays (e.g., see Erickson, 1978). Results demonstrate that increasing the magnification of the TV viewing system, thereby increasing the dimensions of targets at the display and the number of active scan lines making them up, decreases the amount of time required to detect them. Magnification also improves the rate of target recognition across the somewhat restricted range of values tested (i.e., 1.0X to 4.0X). In general, up to size limits that exceed the dimensions of conventional stereo TV display screens, the larger the image of an object on the retinas of an observer, the more defined it is in terms of pixels or scan lines, and the more quickly and accurately it will be recognized by an observer.

Magnification, however, narrows the visual angle in the remote scene which can be

imaged at any given moment. It also places increased demands for precision on the means by which cameras are aligned and aimed. Any inaccuracies in aligning cameras are exaggerated by magnification. This is particularly serious in the case of vertical misalignments in stereo TV images where even small (i.e., < 10 arcmin) misalignments are known to result in discomfort and perceptual distortions. Additionally, any unintended movement or tremors in the remote camera configuration are exaggerated and thereby made more apparent and distracting to an observer. It should be kept in mind that the present line of investigation avoided such difficulties by using only pre-selected, well-aligned static pictures that sampled performance over only a moderate range of magnification values. The success of magnification strategies on actively scanning camera heads, particularly those which have their movements coupled to the head and upper body motions of a remote operator, will depend greatly on the elimination of misalignment problems and the cancelling out of unintended movements.

The effects of wide camera separations on performance were less clear-cut. Performance appeared to be optimized at an intermediate camera separation (i.e., at 6I(381mm)), and it fell off with increases in camera separation beyond this level. The effect was more apparent in the target detection time measure, but it was also found for recognition rate. At this point, these results remain somewhat puzzling, and therefore require replication and further experimental scrutiny before firm conclusions are justified. Results do, however, suggest that the optimal camera separation for detection and recognition of targets at the ranges tested in this experiment is one that deliberately exaggerates disparity cues to the depth and distance of objects.

The significant interactive effect between camera interaxial separation and magnification that was found for target recognition rate suggests that at the higher levels of magnification (i.e., 2.0X and 4.0X) tested, the pattern of increasing efficiency for wider camera separations generally held. However, at the lower magnification levels (i.e., 1.0X and 1.25X), recognition rates were generally much poorer and performance across the varying levels of camera interaxial separation was less stable. This pattern of results does not, however, confirm the impressions of stereo photographers and cinematographers (e.g., McAdam, 1954) who claim that perceptually confusing distortions of space perception result from non-orthoscopic combinations of camera interaxial separation and magnification. At very least, if such distortions do occur, they do not appear to have the predicted disruptive consequences for human target detection and recognition.

References

Cole, R.E., Pepper, R.L., and Pinz, B.L. The influence of head movement parallax on perceptual performance under direct- and TV-displayed conditions. San Diego, CA: NOSC Technical Report No. 678, May 1981.

Erickson, R.A. Line criteria in target acquisition with television. Human Factors, 1978, 20, 573-588.

McAdam, D.L. Stereoscopic perceptions of size, shape, distance, and direction. Journal of the Society of Motion Picture and Television Engineers, 1954, 62, 271-293.

Pepper, R.L., Cole, R.E., and Spain, E.H. The influence of camera separation and head movement on perceptual performance under direct and TV-displayed conditions. Proceedings of the Society for Information Display, 1983, 24, 73-80.

Pepper, R.L., Smith, D.C., and Cole, R.E. Stereo TV improves operator performance under degraded visibility conditions. Optical Engineering, Jul/Aug 1981, 20(4), 579-585.

Spain, E.H. and Cole, R.E. Camera separation, head motion camera coupling, and perceptual performance with a helmet-mounted stereo TV display. Kailua, HI: SEACO, Inc. Report No. 82-11-03, November 1982.

Spain, E.H., Cole, R.E., and Pepper, R.L. Visual performance and fatigue with stereoscopic television displays. Proceedings of the American Nuclear Society National Topical Meeting on Robotics and Remote Handling in Hostile Environments, Gatlinburg, TN, April 1984, pp. 61-67.

Spatial Frequency and Orientation Information
as a Predictor of Discriminability in Symbologies

by

William P. Marshak and John C. Osarczuk

Abstract

A method for describing symbologies and estimating their discriminability is proposed based on the two dimensional fourier transform of the symbols. A signal detection experiment was conducted to test the principles using simple stimuli whose fourier transforms are understood. Increasing differences between target and noise in spatial frequency content and orientation of frequency information lead to greater sensitivity to the target's presence. These results support the proposed use of fourier transforms in describing symbologies.

Introduction

Design of visual display symbology is black art that draws upon tradition, general recommendations (such as McCormick and Sanders, 1982), and experimental comparisons between proposed sets. Human engineers do not have an effective descriptive metric drawn for comparing visual stimuli. Such descriptive methods would permit quantitative comparison of different symbologies and prediction of such key issues as their discriminability.

Classical experiments such as Hitt (1961) seek to behaviorally determine the relative effectiveness of different types of symbols. Yet, Hitt does not compare the performance differences with obvious differences in size, shape and contrast of the stimuli. Green and Pew (1973) come closer to examining the perception issues in symbology by developing a confusion matrix to show substitution errors. They did not tie these errors to stimulus attributes. No one has related behavior measures with stimulus attributes of symbologies. The current discussion will center on a comprehensive theory of symbology description and how it might lead to predicting performance of a symbol set. It will ignore cognitive issues relating to the meanings of those symbols.

Description and comparison of complex symbols can be based on their two dimensional (2-D) fourier transform (Campbell and Maffei, 1975; Ginsberg, 1978). Any symbol can be considered to be a wave form consisting of changing contrast over space. Convolving this wave form into the frequency domain, you can create the two dimensional amplitude or power spectra. This spectra shows the distribution of spatial frequency information based on frequency (lower frequencies near the center, higher frequencies further peripheral) and the orientation along which the frequency information lies. Such a 2-D transform allows concise mathematical description of the distribution of contrast in any

two dimensional image.

How could the 2-D transform be used as a basis for predicting such performance measures as discriminability of symbols? Consider a simple signal detection experiment. Noise stimuli fill a two dimensional display area. These noise stimuli are parallel lines, whose fourier transform has most of its energy along the orientation at right angles to the line's orientation and whose spatial frequences lie at odd harmonics of the fundamental frequency corresponding to line separation. The greater the difference between noise and target, the greater the discriminability of stimuli and the greater the sensitivity of observers to their presence.

Target stimuli with known differences in frequency and orientation of frequencies can be constructed. Parallel lines with different line separations, or rotated in different orientations would permit precise control of the difference between signal and noise. Observer sensitivity should be directly proportional to differences in the 2-D transform for changes in spatial frequency and the orientation those frequencies lie along.

Methods

Subjects were 40 cadet volunteers at the United States Air Force Academy. Their age varied from 18 to 23, all had normal vision or wore correction to 5/5. Each subject was read a short instruction set and was seated at a Zenith Z-120 microcomputer equipped with a 12 inch screen using P-31 phosphor cathode ray tube display. The entire experiment was controlled by the Z-120 using custom software. Further instructions were read from the screen, including examples of target and noise stimuli. The instructions stressed speed and accuracy as well as the independence of trials. No information was given about the proportion trials containing targets (50%) nor was there any attempt to convey a response criterion.

Background light at screen surface was measured at 8.5 cd/meter squared. A typical screen full of stimuli put out 17.9 cd at the approximate viewing distance of 53 centimeters. Viewing distance was not controlled; subjects were allowed to use a comfortable distance within reach of the keyboard for responses. Stimulus contrast was estimated to be approximately 47 percent.

Pixel placement was controlled by the Z-120 character generator and two adjacent characters were used for each stimulus. Vertical pixel pitch was .69 mm and horizontal pitch was .30 mm. Noise stimuli consisted of vertical parallel lines made up of 7 pixels for an overall length of 4.82 mm (4.48 degrees of visual angle) and a horizontal separation of 5 pixels or 1.5 mm (1.4 degrees). A total of 50 pairs of parallel lines were randomly placed in screen memory (out of a possible 24 x 40 = 960 positions) while the screen was blanked.

On signal trials, one additional stimulus was added in an random open screen position. This stimulus consisted of two parallel vertical line pairs, 45

degree slanted line pairs or horizontal lines pairs. The target stimuli consisted of two parallel lines the same length as noise with separations of 2.7 mm or 3.9 mm (2.5 or 3.6 degrees). For slanted lines the inter-pixel distance was .75 mm, and three line separation distances of 1.5, 2.7 or 4.5 mm (1.4,2.5 and 3.6 degrees) were used. Total length of slanted lines was 6 pixels or 4.51 mm. Finally, horizontal lines had an inter-pixel distance of .6 mm (every other horizontal dot), a length of 4.2 mm and the interline distances 1.38, 2.76 or 4.14 mm. These were as close to comparable distances that the Zenith's screen resolution would permit for targets stimuli.

Stimuli took four seconds to build on the blank screen. Upon completion of their construction, the screen was presented during the course of one scan (.0167 seconds) and was maintained for 1.25 seconds. At the completion of the viewing period, the screen was blanked and decision time clock begun. When the subject pressed a key designating signal+noise or noise, the decision clock was read and the data of response and time stored. A four second intertrial interval occurred followed immediately by construction of the next stimulus.

Twenty trials in random order were conducted in this manner. Half the trials contained target stimuli; the other half were noise alone. Upon completion of the trials in approximately 3.33 minutes, the computer printed the performance for each trial and computed the mean decision time and the d' sensitivity measure.

Results

Two measures of performance were used to determine discriminability of the target from background stimuli. The first measure was the parametric version of d' (Green and Swets, 1966). The statistic was computed by formula 1:

$$d' = .707 \times \{z[p(\text{hit})] - z[p(\text{false alarm})]\}$$

(1) } The second performance measure was decision time. Decision time was the time interval between the completion of the 1.25 second observation time until the observer made his or her decision. It differed from reaction time since the observer may have seen the target earlier but could not respond until the display disappeared.

One cell of the analysis, the one in which target and background were identical, did not use subject data. This was an impossible task and would have been frustrating and a waste of subject time. To permit a complete analysis of the variables, data for d' was generated by a monte carlo simulation of a subject's performance on the experiment. A random number generator simulated subject performance and the results were used as the d' data for that condition. The decision time data created a real problem for the identical target and background condition. These data represent random selection from adjacent cell decision times. The d' data is a satisfactory

solution to simulating subject performance. The decision times generated are less satisfactory, but are probably a conservative estimate of the decision times when no criteria were available to the subjects.

The d' results are represented in figure 1. Sensitivity increased with larger differences in the spatial frequency content of the target [Fcalc= 3.5349, Fcrit(df=2,36;alpha=.05)= 3.26]. Sensitivity also increased with larger differences in the orientation of the spatial energies [Fcalc= 18.86, Fcrit(df=2,36;alpha=.05)= 3.26]. There was no interaction between frequency and orientation [Fcalc= .4503].

Decision times are plotted in figure 2. Decision times were shorter with differences in the orientation of spatial frequencies [Fcalc= 5.504, Fcrit(df=2,36;alpha=.05)=3.26]. Decision times were not signifcantly different for either spatial frequency differences [Fcalc= 1.626] or the interaction of frequency and orientation [Fcalc= .508].

Discussion

There is nothing surprising from the finding that greater differences between target and noise signals produce greater sensitivity in receivers and shorter decision time. The real issue was whether frequency and orientation of spatial energy constitue good predictors of differences. These data seem to support this hypothesis. Greater differences along either dimension produce greater sensitivity. Failure to find this effect in decision times may have been a result of the identical target and noise condition and the way the data was fabricated. The data probably grossly underestimates the decision time when possible criteria is lacking.

Another interesting outcome was the lack of an interaction between the frequency and orientation of spatial information. This outcome indicates either independent channels process each attribute or the same channel handles both in a linear fashion.

A practical problem prevents using fourier analysis as a criteria for designing complex discriminable symbologies. The two-dimensional spectra, either amplitude or power, is not a convenient means for comparison. We are working on metrics which could summarize the fourier transform so that comparison of symbols and prediction of their discriminability could be easier.

Bibliography

Campbell, F.W. and Maffei, L. (1975). Contrast and spatial frequency Scientific American, 231(5), 106-114.

Ginsberg, A. (1973). Visual information processing based on spatial filters

constrained by biological data, AMRL-TR-78-129, Vol 1&2.

Green, D.H. and Swets, J.A. (1966). Signal Detection Theory and Psychophysics, New York: John Wiley and Sons.

Green, P. and Pew, R.W. (1978). Evaluating pictographic symbols: An automotive application. Human Factors, 20(1), 103-114.

Hitt, W.D. (1961). An evaluation of five different coding methods. Human Factors, 3(2), 120-130.

McCormick, E.J. and Sanders, M.S. (1982). Human Factors in Engineering and Design, New York: McGraw-Hill.

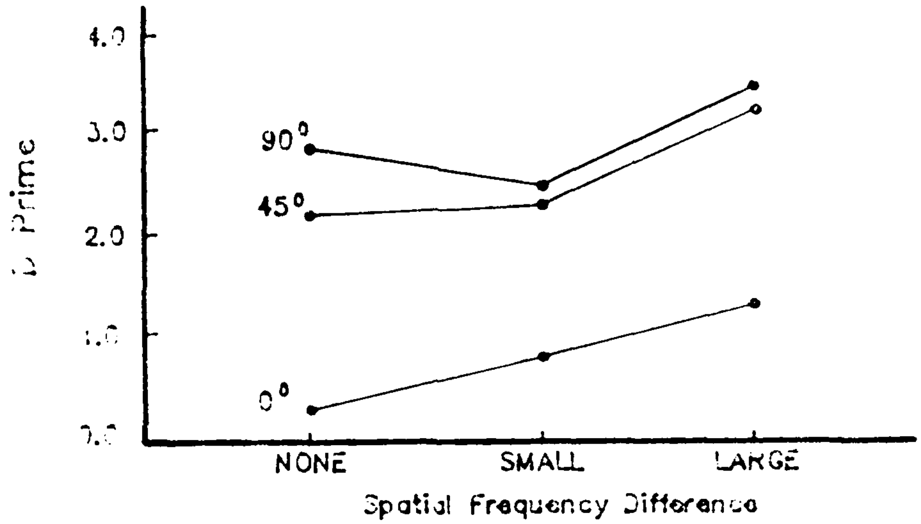


FIGURE 1: Sensitivity as a function of target spatial frequency and orientation

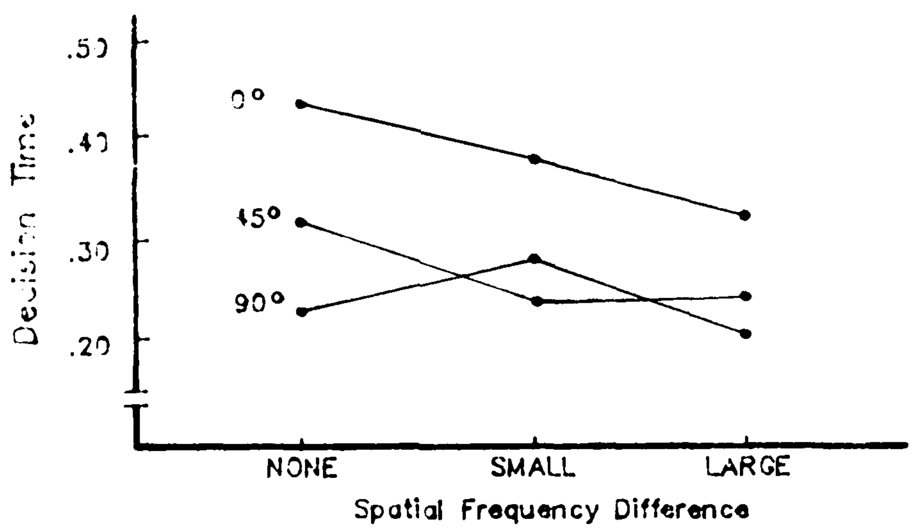


FIGURE 2: Decision time as a function of target spatial frequency and orientation

The Role of Central Field Stimulation in the Perception of Self-Motion: Implications for Flight Simulation

George J. Andersen
University of Illinois, Urbana-Champaign

Abstract

An important consideration for some types of flight simulation is that sufficient visual information be provided for a perception of self-motion. A general conclusion of previous research is that peripheral stimulation (outside a 30° diameter area of the central field) is necessary for perceived self-motion to occur.

Research will be presented that demonstrates that peripheral involvement is not essential to perceive self-motion. Subjects viewed computer-generated displays simulating observer motion through a three-dimensional volume filled with randomly positioned points. It was found that: (1) stimulation of as little as 7.5° of the central visual field can result in a perception of self-motion; (2) perceived self-motion may be related to the perceived depth of the display; (3) greater postural instability occurred with displays simulating backward motion rather that forward motion; and (4) stimulation of the central field can result in motion sickness. The implications of this research for flight simulation will be discussed.

In some cases of flight simulation training, such as low altitude flight or landing, it is critical that sufficient information be available to the pilot to determine the speed of his motion and the distance to the ground. Lack of sufficient information may result in misperceived speed and distance which could result in poor training performance. In order to determine correctly the speed of the pilot and the distance to the ground, the pilot must correctly perceive his own motion (or self-motion).

A general conclusion of previous research on the perception of self-motion is that peripheral stimulation is necessary for a visually-induced perception of self-motion to occur (Brandt, Dichgans, & Koenig, 1973; Held, Dichgans, & Bauer, 1975). More specifically, it has been concluded that stimulation outside a 30° diameter area of the central visual field is necessary for a perception of self-motion. Andersen and Braunstein (1985) recently found that linear vection can be induced with a radially expanding flow pattern presented to central regions of the visual field. The stimuli were computer-generated simulations of forward motion through a volume uniformly filled with randomly positioned points.

Three variables were investigated: the visual angle of the display, the density of the dot pattern, and the simulated speed of observer motion. Induced self-motion was reported at all levels of visual angle ranging, from 7.5° to 21.2°. The duration of induced self-motion was measured by having the subjects press a button whenever they perceived that they were moving. A significant interaction was found for the duration of reported self-motion between the simulated speed of

the observer and the visual angle of the display. The duration of reported self-motion was smallest for the largest visual angle examined (21.2°) at the highest speed conditions. These results contradicted the conclusion of other investigators (Brandt et al., 1973; Held, et al., 1975) that peripheral stimulation is essential for induced self-motion.

In a second experiment, Andersen and Braunstein asked subjects to rate their impression of depth for the same stimuli under the same viewing conditions. The interaction of speed of observer motion and visual angle found in the first experiment was also found in the second experiment, with the highest depth ratings occurring in the conditions in which the longest durations of self-motion had been reported. Andersen and Braunstein suggested that these results are indirect evidence that induced self-motion from stimulation of the central visual field is dependent on internal depth within the display. They concluded that the failure of previous research to find self-motion effects in central vision was due to the lack of optical flow information for internal depth in the stimuli that were investigated.

The mechanisms involved with determining perceived self-motion and spatial orientation are also important in the control of postural stability. Lee and Lishman (1975) placed subjects within a large enclosure where the walls were not attached to the floor and moved the enclosure back and forth while the subject remained stationary. They found that subjects placed in this environment exhibited great difficulty in maintaining balance. Similar results were found when infants were placed in the enclosure (Lee & Aronson, 1974). The visual information resulting in postural instability in these studies completely surrounded the subject horizontally.

The purpose of the experiment that I'll be reporting today was to examine the effects of optical flow in the central field on postural stability. In our earlier research, we found that induced self-motion occurred when stimulation was limited to a small area of the central visual field. Lee and Lishman (1975) indirectly examined the effects of central stimulation on postural stability. They found that postural instability was reduced when a narrowly structured object was placed in front of the subject while standing in the moving chamber. However, their study involved the removal of information in the central field and did not examine the importance of the type of flow pattern in the central field.

In the present experiment, subjects viewed four different optical flow patterns which simulated four different observer motions: forward motion, backward motion, forward motion with a vertical displacement, or backward motion with a vertical displacement. A platform with a fulcrum point in the center was used to measure body sway. Three different visual angles were examined. The subjects were instructed to keep the platform as level as possible while viewing the displays.

Method

Subjects. The subjects were four students in lower division psychology courses at the University of California, Irvine, who received extra credit for their participation. Vision of at least 20/40 (Snellen Eye Chart) was required in both eyes.

Design. Three independent variables were examined: The visual angle of the display (7.5, 10.6, or 15°), the simulated direction of observer motion (forward or backward) and the presence or absence of a vertical sinusoidal displacement. All variables were run as within subjects variables.

Stimuli. The stimuli were two 100-s (2400-frame) computer-generated 16-mm motion picture sequences. The displays were generated one frame at a time on a Tektronix 4010 display scope, recorded on Kodak Plus-X film, and printed on high contrast film. Each display simulated motion of an observer through an area uniformly filled with randomly positioned points. The dimension of the space was 12 by 12 by 18 units, with the projection plane 3 units from the simulated viewpoint of the observer. In the display with the sinusoidal displacement, the dots were displaced along the vertical axis according to the function $Y = 2\cos(\pi)\cos(3\pi)$. Whenever a dot was projected outside the space by the vertical displacement it was replaced at the opposite end of the space along the vertical axis at the same X and Z coordinates.

Apparatus. A 16-mm projector was used to present the film. The film was projected through a glass window onto a first surface mirror and onto a translucent screen. The subject viewed the displays while standing in a large moveable booth. A platform 61cm X 61cm was used to measure the subject's postural stability. Four linear potentiometers (Allied Electronics Type M 1326-1) were used to measure displacements of the right, back, left, and front edges of the platform. Each potentiometer had a maximum displacement of 1.27 cm. Tests conducted on the potentiometers indicated that the third and fourth potentiometers were the most accurate in representing forward-backward and left-right displacements of the platform, respectively. Thus these two potentiometers were used in the analysis.

Results

The output of each potentiometer was sampled at a rate of 100 samples per second. From these samples the variance of the potentiometer output was computed for each display. The effects of the three variables (visual angle, direction of motion, and vertical displacement) were examined in an analysis of variance for the two potentiometers. The only significant results were found for the third potentiometer, which was located behind the subject. The main effect for direction of simulated motion was significant $F(1,3) = 44.94$, $p < .01$, with a greater amount of variability occurring for the displays simulating backward motion. The interaction of simulated direction and vertical displacement $F(1,3) = 12.7$, was also significant $p < .05$. The least variance occurred for the simulated forward display when compared to the other three displays. The main effects for visual angle $F(2,6) < 1$ and vertical displacement $F(1,3) < 1$ were not significant and there were no other significant interactions.

Discussion

These results demonstrate that postural stability can be affected by information limited to the central visual field. Taken together with the results obtained by Andersen and Braunstein (1985), and corroborating findings recently published by Stoffregen (1985), these results suggest that the central visual field

contributes to mechanisms responsible for spatial orientation and self-motion. An important question to consider is what type of mechanisms might be responsible for a perception of self-motion. It is proposed that the following three mechanisms are involved in the perception of self-motion from radially expanding flow patterns. (See Table 1.) These mechanisms would be based on the types of perceptual processors discussed by Hoffman and Bennett (in press). According to their theory, processors concerned with the interpretation of an ambiguous retinal image must take advantage of regularities of the world as environmental constraints in order to determine unique solutions. The first processor would determine that each point in the display is moving in depth. Consider the motion of each dot in the pattern of optical flow. The projected velocity function of each dot is exponential. If the visual system exploits the regularity that objects tend to move at constant three-dimensional velocities, then the exponential velocity function in the two-dimensional projection will result in a perception of motion in depth. The second processor would be concerned with determining the relative distance among the points as well as the rigid three-dimensional structure of the environment. For an observer whose line of sight is parallel to the (observer's) direction of motion, the ratio of the projected velocity of a dot to its projected radial position from the focus of expansion is constant. However, this ratio will vary among dots in different depth planes. If the visual system exploits the regularity that objects move at the same three-dimensional velocities relative to the observer, then these variations in the ratio of velocity to radial distance will be perceived as variations in depth within a rigid structure. The third processor would be concerned with determining a perception of observer motion. If a common motion vector is extracted from the three-dimensional velocity field of objects that appear to be independent, then the constraint may be exploited that independent objects do not move in the same direction and at the same three-dimensional velocity unless the observer is moving.

Table 1
Examples of Processors Useful in the Perception of Observer Motion.

	INFORMATION IN OPTIC FLOW	REGULARITY	OUTPUT
PROCESSOR 1	2-D Exponential Motion	Objects Move at Constant 3-D Velocities	Motion in Depth of a Single Dot
PROCESSOR 2	Ratios of 2-D Velocities to Projected Radial Distance	Objects Move at the Same 3-D Velocities	Relative Depth Rigid Structure
PROCESSOR 3	Common Motion in 3-D Velocity Field	Common Motion of Independent Objects Unlikely	Observer Motion

The results of this research indicate that stimulation of the central visual field is sufficient for determining spatial orientation. This suggests that for some types of flight simulation, especially simulation of low altitude flight or landing, optical flow information within the central visual field is important. In particular, relative depth information may be critical for determining spatial orientation from this area of stimulation. This suggests that the addition of relative depth information within the pilot's immediate area of fixation may be critical for effective training in these types of simulation. Failure to present this type of information in the immediate area of fixation could result in poor training effectiveness and low transfer of training to real world flight.

References

Andersen, G. J., & Braunstein, M. L. (1985). Induced self-motion in central vision *Journal of Experimental Psychology: Human Perception and Performance, 11,* 122-132.

Brandt, T., Dichgans, J., & Koenig, E. (1973). Differential effects of central versus peripheral vision on egocentric and exocentric motion perception. *Experimental Brain Research, 6,* 476-491.

Held, R., Dichgans, J., & Bauer, J. (1975). Characteristics of moving visual scenes influencing spatial orientation. *Vision Research, 15,* 337-365.

Hoffman, D., & Bennett, B. (in press). Visual representations: Meaning and truth conditions. In S. Steele and S. Schiffer (Eds.), *The Second Arizona Colloquia on Cognitive Science.* University of Arizona Press.

Lee, D. N., & Aronson, E. (1974). Visual proprioceptive control of standing in human infants. *Perception & Psychophysics, 15,* 529-532.

Lee, D. N., & Lishman, J. R. (1975). Visual proprioceptive control of stance. *Journal of Human Movement Studies, 1,* 87-95.

Stoffregen, T. A. (1985). Flow structure versus retinal location in the optical control of stance. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 554-565.

Image Processing and Information Extraction

Gilbert G. Kuperman
Harry G. Armstrong Aerospace Medical Research Laboratory

Abstract

An imagery interpretability rating scale was applied to spatially filtered versions of digital images of tactical aircraft. The spatial filters employed were based on a model of visual psychophysics. A regression analysis was performed to relate the center frequency of the filter to the achieved rating scale value. Good agreement ($r^2 = 0.56$) was found using the center frequency as the predictor variable.

Introduction

Numerous researchers have attempted to apply linear systems theory constructs to models of visual perception. Carel, Herman, and Olzak (1978), for example, present a family of modulation sensitivity function curves which depict the modulation required by an observer, as a function of spatial frequency, in order to differentiate a sinusoidal luminance grating from an unmodulated field of the same average luminance. Ginsburg (1978) described a similar model, termed the Contrast Sensitivity Function (CSF), which decomposed the visual describing system function into a set of bandpass spatial filters. He posited that the filters were Gaussian in shape, were separated from each other in center frequency at one octave intervals, and exhibited a bandwidth of one to two octaves.

Many image quality assessment measures have been developed and applied. Kuperman (1980), for example, compares several subjective and objective measures using imagery degraded by induced motion. One type of approach has been based on information extraction tasks supported by different image quality levels. AIR STD 101/11 (1978) is an image interpretability rating scale which assigns rating levels as a function of the amount of detail that can be extracted from an image.

The present research is an attempt to relate the bandpass spatial filtering model of visual perception and the imagery interpretability rating scale. The underlying assumption was that the individual filters should support distinct levels of interpretability.

Methodology

Stimuli

Photographs of six fighter aircraft were digitized as 256 X 256 picture elements (pixels) X 8 bits of intensity. The digital representations were subjected to a two-dimensional, discrete Fourier transformation. The Fourier Transform decomposes the image into sinusoidal components of integer fundamental frequency which is expressed in terms of cycles/array dimension. Seven, bandpass filtered, versions of each aircraft were created. The center frequencies of the filters were 0.25, 0.5, 1.0, 2.0, 4.0, 8.0, and 16.0 cycles/object dimension. The length and wingspan of each aircraft were measured, in pixels, and their average was used as the object dimension in relating the spatial filters to the aircraft dimension. The filters were Gaussian-weighted and had individual bandpasses of 1.5 octaves (with respect to the center frequency). The images were then inverse transformed and reconstructed as positive photographic prints. (The digital image processing facility employed in creating the stimuli is described in Kuperman, Wallquist, and Katz, 1984.) A total of 42 prints were created (six aircraft X seven filters).

Subjects

Seven Air Force image analysts, three of whom were trained imagery interpreters, served as subjects (Ss). All were male and had 20/20 visual acuity (corrected of uncorrected). All were highly experienced (minimum 12 years) and all were familiar with the ASCC rating scale.

Practice

A set of practice images (of a white square on a black background) was processed through the passband filters. The Ss' task was to rate the level of information (similar to the ASCC scale) that could be extracted from each of the seven resultant images. In addition, the Ss were required to state their confidence in their rating. The practice trials were to familarize the Ss with the confidence rating procedure and with the mechanics of the task.

Task

The experiment was self-administered by the Ss. Each S first accomplished the practice trials. The S then rated the 42

experimental stimuli for both interpretability (using the ASCC rating scale) and confidence in reported interpretability level. Each S was presented with the experimental stimuli in a unique, random sequence.

Analyses and Results

Interpretability Ratings

The interpretability ratings were converted into ranked data. A single factor analysis of variance (ANOVA) was performed to test for differences between aircraft types. No significant difference was found and the data set was collapsed across this variable.

A two factor ANOVA was then performed on the ranked interpretability data (pooled across aircraft types) for the main effects of Filters (F), Subjects, and the F X S interaction. The seven filtered image versions were found to be highly significantly different from each other ($p < 0.01$) with regard to the interpretability ratings that they produced. Ss were found to be significantly different from each other ($p < 0.05$). (The F X S interaction was not found to be significant.)

A *post hoc* computation, the omega^2 statistic, was performed on the ranked interpretability data to determine the degree to which each factor in the design contributed to the total variance encountered in the resultant data. The main effect of Filters accounted for 84 percent of the total variance while that of Ss accounted for only five percent. (The F X S interaction, while not significant, contributed six percent of the total observed variability.)

A *post hoc* test, Tukey's HSD statistic, was employed to investigate the significant difference between interpretability ratings produced by the seven Filters. No difference was found between the two Filters with the lowest center frequencies (0.25 and 0.5 cycles/object dimension). Every other Filter was different ($p < 0.05$) from these two (as a group) and from each other.

The main effect of Ss was also subjected to analysis by Tukey's HSD. No systematic pattern was found between the age or experience of the Ss which accounted for the differences in producing interpretability ratings.

A two factor ANOVA was also performed for rating confidence. The main effect of F was found to be significant ($p < 0.05$). In general, confidence was highest for filters with the lowest and highest center frequencies.

Regression Analysis

A regression analysis was performed to relate the center frequencies of the bandpass filter to the ASCC interpretability ratings produced by the filtered images. The regression equation that resulted was:

$$\text{Ratings} = 0.431\ (\text{Center Frequency}) + 1.579$$

where the Center Frequency was expressed in cycles/object dimension. The r^2 for this equation was found to be 0.562.

Conclusions

The individual spatial filters which, together, compose the CSF, demonstrated a relationship to ratings produced using an image interpretability scale. Since the rating scale was based on distinct levels of information extraction tasks that would be supported at each rating level, it is reasonable to infer that the spatial filters also correspond to distinct levels of information extraction.

Several areas of application suggest themselves. The spatial filters could form the basis for the design of image compression algorithms which are optimized for the information requirements of the human visual system. Standardized filtered imagery sets could be employed in the assessment of display image quality (softcopy) or in the evaluation of imagery interpretation equipment (hardcopy).

References

Air Standardization Coordinating Committee (ASCC), Imagery Interpretability Rating Scale, AIR STD 101/11.

Carel, W. L., Herman, J. A., and Olzak, L. A., Design Criteria for Imaging Sensor Displays, ONR-CR213-107-1F, Office of Naval Research, May 1978.

Ginsburg, A. P., Visual Information Processing Baased on Spatial Filters Constrained by Biological Data, AMRL-TR-78-129, Aerospace Medical Research Laboratory, December 1978.

Kuperman, G. G., Investigation of Photographic Image Quality Estimators, AFAMRL-TR-80-27, Air Force Aerospace Medical Research Laboratory, April 1980.

Kuperman, G. G., Wallquist, D. L., and Katz, L., "A Digital Image Processing Facility for Human Factors Research," Proceedings of the Human Factors Society 28th Annual Meeting, October 1984.

Kuperman, G. G., Bandpass Spatial Filtering and Information Content, AAMRL-TR-85-046, Harry G. Armstrong Aerospace Medical Research Laboratory, July 1985.

Visual Evoked Potentials to Luminance and Chromatic Contrast
During Normal Viewing and Following Laser Flash Exposure

Fred H. Previc, Douglas J. Coffey and Michael F. Blankenstein
Technology Incorporated

Abstract

The present study compared the spatiotemporal tuning characteristics of the luminance (achromatic) and chromatic systems, as well as their vulnerability to intense laser flashes. Visual evoked potentials (VEPs) were recorded from chronically implanted electrodes in two rhesus monkeys in response to counterphasing luminance and chromatic (red-green) sine-wave gratings. The results indicated that the chromatic system exhibits lowpass tuning in both the spatial and temporal domains, and that, relative to the luminance system, it is especially poor at resolving high spatial frequencies. Also, the effects of intense 100-msec laser flashes at 476.5, 514.5 and 647.5 nm were generally greater for the chromatic VEPs, even when the spectral content of the flash and luminance grating were more similar. These findings have implications for the use of color contrast in cockpit displays.

Introduction

Information in the visual environment consists of two types of contrast -- *achromatic (luminance) and chromatic* (color). Achromatic contrast is defined as luminance variations in a stimulus, whereas chromatic contrast refers to differences in spectral content among stimuli of equal luminance. There has been an increasing use of color contrast in the cockpits of modern aircraft. While the use of color presents several advantages over information presented using black-white or other forms of luminance contrast, it may possess some disadvantages as well. The present study attempted to explore some of the problems associated with the use of color contrast, both under normal viewing conditions and under military visual threat conditions -- in this case, exposure to intense laser flashes.

Many investigators, primarily using psychophysical techniques, have attempted to compare the spatial and temporal response characteristics of the luminance and chromatic contrast systems. Whereas the luminance system appears, at least under most conditions, to possess "bandpass" properties (i.e., is tuned to intermediate spatial and temporal frequencies) (Kelly, 1977), the chromatic system exhibits "lowpass" tuning, in that it responds best to information presented at low spatial and temporal frequencies (Kelly, 1983). This indicates that color perception may have evolved in the primate in order to discriminate slowly moving or stationary objects (such as fruits) at a close distance, and may not be particularly useful in the perception of detailed, rapidly changing information.

Another important aspect of chromatic contrast perception concerns its vulnerability to intense light adaptation. Some evidence, primarily neurophysiological, suggests that chromatic contrast perception may be less

impaired following exposure to intense light of one color because the remaining color mechanism may be left undisturbed (De Valois, 1965). Other evidence indicates that color-contrast perception (as assessed using equiluminant red-green flicker) may be more impaired than luminance-contrast perception, at least for certain adapting wavelengths (Nagy & Purl, 1985; Varner, Jameson, & Hurvich, 1984). Additional evidence suggests that color contrast perception is more rapidly degraded under stabilized viewing conditions, which produce a substantial amount of neural adaptation (Kelly, 1983).

The purpose of the present study was a two-fold one. Using the visual evoked potential (VEP) as a measure of visual function, the spatiotemporal response characteristics of the red-green chromatic system were compared to those of its achromatic counterpart under normal viewing conditions. In addition, the recovery of VEPs to red-green chromatic gratings following intense laser flash exposure was compared to that of VEPs to red-black, green-black and blue-black luminance gratings.

Method

Two adult rhesus monkeys with normal vision served as subjects in these experiments. Each had several bipolar electrodes chronically implanted in its primary visual cortex (area 17). VEPs reflected localized visual activity which occurred between the electrode's two leads, one of which rested on the cortical surface and the other of which lay 3 mm beneath in the white matter. During each recording session, the animal was paralyzed using Flaxedil and anesthetized by means of Nembutal. VEPs were recorded using monocular stimulation, with a contact lens placed on the right eye to refract the animal and prevent corneal desiccation.

Steady-state VEPs were recorded in response to counterphasing sine-wave gratings which were produced using either luminance or chromatic contrast. In generating the chromatic gratings, the equiluminant outputs of the red and green phosphors of a color CRT were modulated in antiphase; conversely, the luminance gratings were generated either using in-phase modulation of the red and green phosphors (so as to produce high-contrast yellow-black gratings) or by modulating only one of the three CRT channels (so as to produce red-black, green-black and blue-black gratings). In generating the spatiotemporal tuning functions for the chromatic and luminance gratings, four spatial frequencies (1, 4, 8 and 12 c/deg, all modulated at 3 Hz) and four temporal frequencies (3, 6, 10 and 15 Hz at 1 c/deg) were employed. In the laser flash experiment, a 1-c/deg grating modulated at 3 Hz was used as the target stimulus, since preflash VEPs to the chromatic and luminance gratings were of comparable amplitude for this stimulus. All gratings were presented in a 6-deg circular display, at a viewing distance of 2 m. The average luminance of the stimuli was 15 cd/m^2 for the red-green and yellow-black gratings, and 7.5 cd/m^2 for the other gratings.

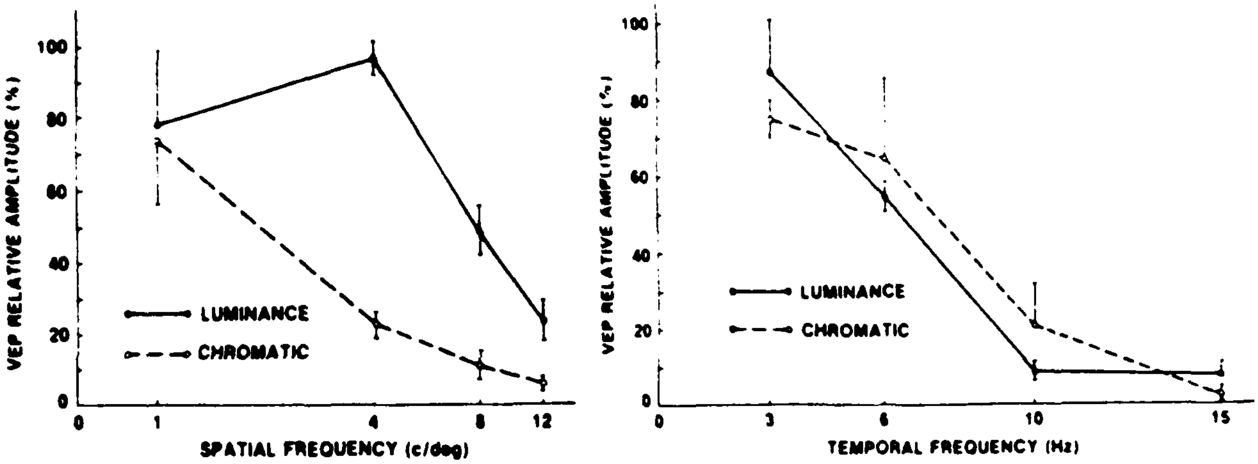
In the laser flash condition, the 476.5-nm (blue) and 514.5-nm (green) outputs of a Spectra-Physics argon 170 CW laser and the 647.5-nm (red) output of a Spectra-Physics 171 krypton CW laser were used. Each laser

exposure consisted of a 100-msec flash which was presented at a retinal illuminance of 8.8 log td. The center of the flash, whose diameter subtended approximately 3 deg on the retina, was aligned with the center of each electrode's receptive field.

VEPs were amplified at a gain of 20,000 using filter settings of 1 and 100 Hz. A PDP 11/34 computer was used to digitize each VEP and to perform a Fourier analysis of its waveform. VEP amplitude values were derived from the Fourier power at the stimulus-reversal frequency.

Results

The results of the spatiotemporal tuning experiment are shown in Figures 1a and 1b. In these figures, the peak amplitude for each electrode



(a) (b)

Figure 1

was calculated, and all other amplitude values were expressed as a percentage thereof. The bars indicate the 95% confidence intervals for VEPs from a total of eight electrodes (five in one monkey and three in the other), all of which were situated in the foveal projection region of area 17. As shown in the figures, VEPs to the chromatic grating exhibited lowpass tuning in both the spatial and temporal domains, whereas VEPs to the luminance grating exhibited bandpass spatial tuning and lowpass temporal tuning. The most obvious difference between the luminance and chromatic VEPs was evidenced in the dramatic falloff in the chromatic VEP beyond 1 c/deg.

The effects of the blue, green and red laser flashes upon the chromatic and luminance VEPs are shown in Figure 2. In this figure, VEPs were averaged across five electrodes (three from one monkey and two from the other), four trials, and successive 5-sec intervals throughout each trial. Each graph reflects the following sequence of events throughout the laser

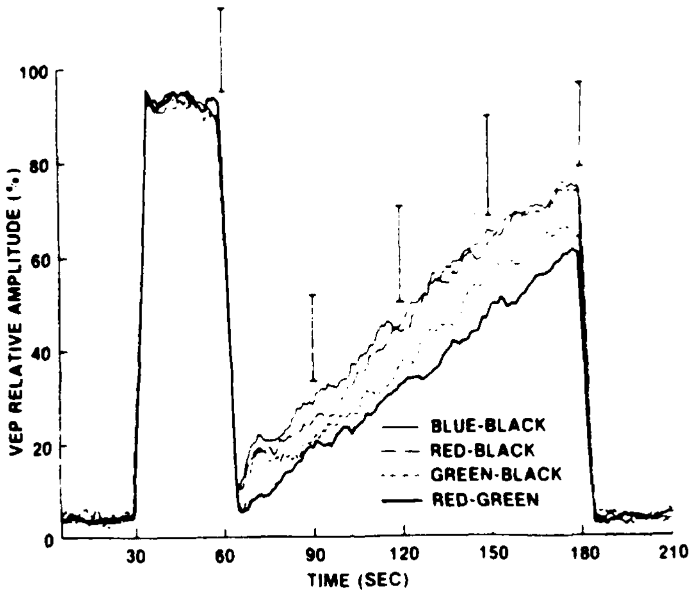


Figure 2

exposure trial: a 30-sec baseline period (during which the monkey viewed a steady, homogenous field), a 30-sec preflash stimulus epoch, the presentation of the laser flash at 60 sec, a 120-sec postflash recovery interval, and a final 30-sec baseline period. Once again, the error bars represent the 95% confidence limits for the VEP at various time periods, based on a standard error estimate averaged across all three laser flashes and all four test gratings. As illustrated in this figure, VEP recovery to the chromatic grating did not progress as rapidly as the recovery to the three luminance gratings. In fact, chromatic VEP recovery was never faster than luminance VEP recovery, even when the luminance grating was the same color as the laser flash (i.e., red flash/red-black grating; blue flash/blue-black grating, etc.).

Discussion

The results of this study suggest that the use of color contrast in the military cockpit possesses at least two distinct disadvantages. First, the visual system cannot resolve detailed spatial information contained in cockpit displays if it is presented exclusively using color contrast. Second, color-contrast perception may be more vulnerable to visual threats such as intense laser flashes in the visible region of the spectrum.

The findings concerning the spatiotemporal tuning of the chromatic and luminance systems are partially consistent with the previous literature. One discrepancy concerns the lowpass temporal tuning of the luminance VEP, which is inconsistent with the bandpass tuning which has been demonstrated psychophysically (Kelly, 1977). In many instances, and especially under

anesthesia (van der Marel, Dagnelie, & Spekreijse, 1981), the VEP has not been shown to exhibit a good high-frequency temporal response, so that this discrepancy may be related to the nature of the VEP itself. On the other hand, the differences between the chromatic and luminance VEPs in their spatial frequency tuning are highly similar to those previously observed both psychophysically and electrophysiologically (Kelly, 1983; Thorell, De Valois, & Albrecht, 1984).

The second important finding was that VEPs to a red-green grating are highly vulnerable to the effects of an intense laser flash, regardless of its wavelength. This finding is consistent with various reports which have suggested that the red-green color-contrast response of the visual system may be more affected by chromatic adaptation and image stabilization than the luminance system (Kelly, 1983; Nagy & Purl, 1985; Varner et al., 1984).

Thus, the findings of this study stress the need for caution in the design and employ of color-contrast CRT displays in the military cockpit. It should be emphasized, however, that a combination of luminance and color contrast, as is typically found, may not suffer from the same disadvantages as an exclusively color-coded display.

References

De Valois, R.L. Behavioral and electrophysiological studies of primate vision. In W.D. Neff (Ed.), Contributions to sensory physiology. New York: Academic, 1965.

Kelly, D.H. Visual contrast sensitivity. Optica Acta, 1977, 24, 107-129.

Kelly, D.H. Spatiotemporal variation of chromatic and achromatic contrast thresholds. Journal of the Optical Society of America, 1983, 73, 742-750.

Nagy, A.L., & Purl, K.F. Adaptation of short wavelength cones: Effects on red-green color thresholds. Investigative Ophthalmology and Visual Science (Suppl.), 1985, 26, 183.

Thorell, L.G., De Valois, R.L., & Albrecht, D.G. Spatial mapping of monkey V1 cells with pure color and luminance stimuli. Vision Research, 1984, 24, 751-769.

van der Marel, H., Dagnelie, G., & Spekreijse, H. Pattern evoked potentials in awake rhesus monkeys. Investigative Ophthalmology and Visual Science, 1981, 21, 457-466.

Varner, D., Jameson, D., & Hurvich, L.M. Temporal sensitivities related to color theory. Journal of the Optical Society of America (A), 1984, 1, 474-481.

Acknowledgment

This research was supported by Contract F33615-84-C-0600 from the USAF School of Aerospace Medicine, Brooks AFB, Texas. The experimental use and care of the animals were in accordance with the Animal Welfare Act of 1970 and the Guide for the Care and Use of Laboratory Animals, prepared by the Institute of Laboratory Animal Resources, National Research Council.

VISUAL EVOKED RESPONSE IN THE PERIPHERY, THE BEGINNINGS OF AN OBJECTIVE MEASURE OF G-INDUCED PLL.

K.M. KENNER
SYNERGY INC. DAYTON, OH.

A.M. JUNKER
AIR FORCE ARMSTRONG AEROSPACE MEDICAL RESEARCH LAB.
WRIGHT-PATTERSON AFB, OH.

R.T. GILL,
DEPARTMENT OF ENGINEERING SCIENCE
UNIVERSITY OF IDAHO, MOSCOW ID.

Work performed at the Air Force Harry G. Armstrong Aerospace Medical Research Laboratory. AFAAMRL/HEG, WPAFB, Ohio

ABSTRACT

High levels of Gz stress can cause aircrew performance degradation, and has been implicated in the loss of life and equipment with high performance aircraft. The danger in head-to-foot acceleration is the loss of cranial blood supply; the first symptom is Peripheral Light Loss (PLL), leading to grey-out (loss of central vision) and eventually to Loss Of Consciousness (LOC). It would be desirable to have an objective, physiologically based, methodology for determining the onset of PLL. A technique has been developed to stimulate the periphery of the visual field with sinusoidally modulated light and to detect a Visual Evoked Response (VER) as measured with the EEG. The VER is at the same frequency as the stimulus. Measurements were taken under three conditions, 1). Baseline, i.e., eyes masked from stimulus, 2). Full stimulus level and 3). Degraded stimulus condition (stimulus power reduced by 10dB, to simulate the onset of PLL). The EEG (Output) and the light stimulus (Input) were measured across time. Fourier analysis was applied to these measures to generate frequency domain values. Measures of both Coherence and RMS Power were recorded for possible correlation with experimental conditions. Results have shown Coherence (an Output/Input ratio that indicates linear cause and effect relationships for sinusoidally stimulated systems) to be a sensitive measure of peripheral stimulation. The sensitivity of this measure indicates the potential for the development of an objective measure for Gz-induced peripheral light loss.

INTRODUCTION

The Air Force has placed an emphasis on performance assesment metrics. There is a strong emphasis on the development of objective, physiologically based, metrics for workload/stress. This is especially important with high performance aircraft, as even a momentary lapse in peak performance can endanger both life and equipment. To date measures of Gz (head-to-foot) acceleration effects have been confined to laboratory simulators and require an active focus of attention by the human operator. These mesures are accomplished by having the operator signal how far, on a light bar, his/her peripheral vision extends. This

method precludes the operator from engaging in mission related tasks, and therefore is unsuitable for in-flight use. An EEG based method of determining the onset of Gz-induced stress would have several advantages; the most notable advantage would be the possibility of in-flight crew monitoring.

By using appropriate signal averaging techniques, it is possible to detect a response in the human EEG to evoking stimuii. When the stimulus is sinusoidally modulated light the result is called a steady state Visual Evoked Response (VER). Research in this area (Spekreijse, 1966; Wilson and O'Donnell, 1981, and Kenner, Junker and Levison, 1985) suggests that the steady state VER may be a useful indicator of internal cortical functioning.

The paradigm for this experiment postulated that: if a steady state Visual Evoked Response could be elicited in the periphery then PLL would cause a reduction in the the evoked response, and should be measurable with the EEG. To simulate PLL, power in the stimulus was reduced during the experimental conditions by 10 dB. Observations were made to determine if the EEG would be sensitive to the change (or lack of change) in the peripheral stimulation.

METHOD

The experimental apparatus consisted of a chair for the subject to sit in, facing a table with a chin rest (Fig. 1). Directly across from the chin rest was a very low constant intensity fluorescent light with an "X" drawn on the center. The "X" served as a fixation point. By having the subject maintain foveal fixation we were assured that the peripheral stimuli fell only upon the periphery. The lamps used for the evoking stimulus were 26 cm fluorescent light tubes, mounted vertically on the presentation table, 18 in. from the center-point of the subject's chin rest. For this experiment the lights were at a 45 degree angle relative to the center line of the subject's line of sight.

Beckman silver/silver chloride electrodes were used with the Grass model P511AC amplifiers, with amplification x50,000 and bandpass of 0.1 to 300 Hz, to record the EEG. The sinusoidally modulated light was generated with an Exact Eletronics model 7060 sine wave generator and an R. Gerbrands Co. light driver. The output of the flickering lights were measured with a United Detector model 10D high speed photocell, placed on the left light. The stimulus was presented at an average intensity of 200 Ft/L and 60% Depth of Modulation measured at the subject's viewing area. The two channels of data (EEG and Photocell) were fed through General Radio low pass filters (cutoff at 50 Hz). The filtered signals were analyzed on a Nicolet 660B Dual Channel FFT Analyzer. Data was recorded at a sample rate of 100 Hz and the computations of Coherence and RMS Power were calculated using eight 2 second windows. Measures of both Coherence (an Output/Input relationship for a sinusoidally stimulated system) and RMS Power were recorded three times per minute at the target frequency.

Each experimental run consisted of: 1). Two minutes BaseLine (BL) i.e. the subjects eyes masked, with the stimulus flickering, 2). Two minutes of eyes open with Full Stimulus power (FS), 3). A 20 sec 'dead spot' where the stimulus either was reduced 10 dB (Change) or remained the same (No Change), with no data collection, and 4). A final 2 min of data collection in the same condition as 3. Each subject was run at 4 frequencies in both the Change and No Change conditions, for a total of 8 runs.

RESULTS and DISCUSSION

It was obvious after only a few runs that RMS Power exhibited too much variability to be a sensitive measure. Coherence, however, demonstrated significant sensitivity and the data seemed to indicate the potential for reliability. The resulting plots of Coherence, for two subjects, are shown in figures 2 and 3. The plots show the two conditions (Change and No Change) for each subject for each of the frequencies. The abscissae are marked in twenty second intervals. Coherence measures are shown as a percent of the input signal that is present in the output (EEG), at the target frequency.

It becomes apparent from examining the plots that this methodology is not yet sufficiently reliable, but there are noteworthy observations to be made from the data. Most notable is the fact that it was possible to elicit a steady state VER in the periphery. Secondly the predicted effect was observed for some of the subjects at certain frequencies (the best example is Sub 07 at 8.75 Hz, Fig. 3).

The fact that subjects could demonstrate the predicted response characteristics at some frequencies but not at others would lend credence to earlier postulations that individuals have specific frequency sensitivity. Furthermore certain subjects have background (i.e. nonevoked) resonance over certain frequency ranges; this background resonance could bias the results as natural activity in these ranges would mask an Evoked Response.

These novel findings may yet pave the way for the development of an objective measure of PLL, but to achieve this there are at least two things necessary: 1). A much faster measure is needed - the recording interval needs to be less than 3 sec, and 2). The need for a more reliable measure. Perhaps just finding the "best" frequency for each subject would resolve the second problem. Even with the above noted shortcomings, this methodology is a first step towards developing an unobtrusive and objective measure of Gz-induced stress.

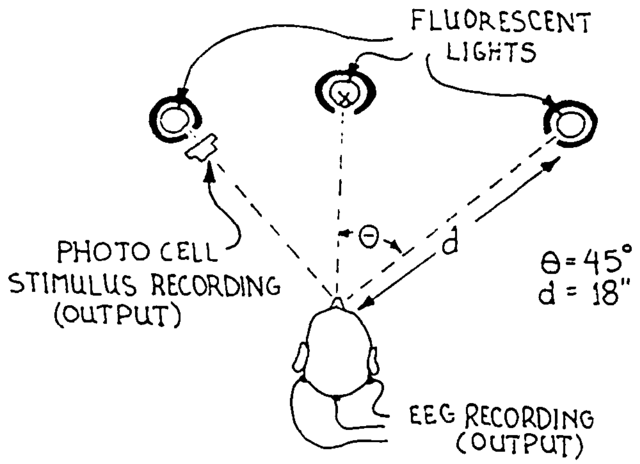


FIGURE 1.
EXPERIMENTAL SET-UP

FIGURE 2.
DATA FOR SUBJECT 05

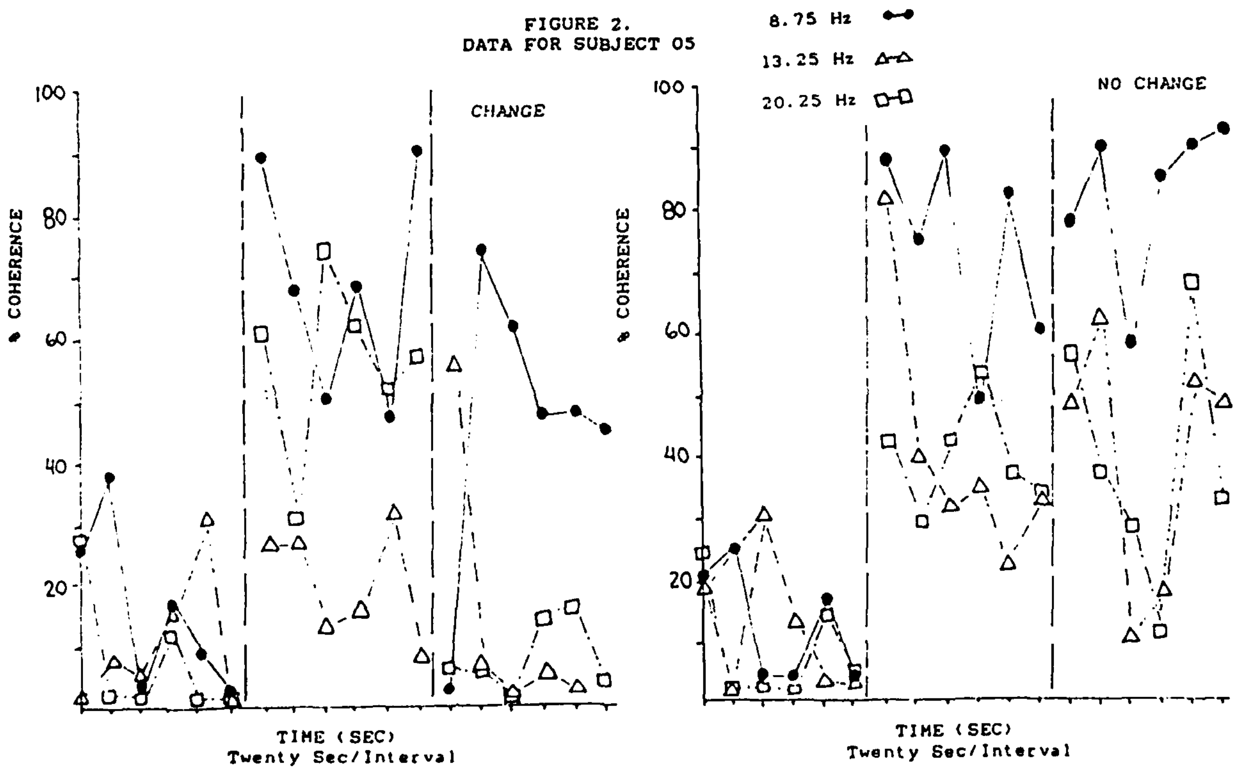
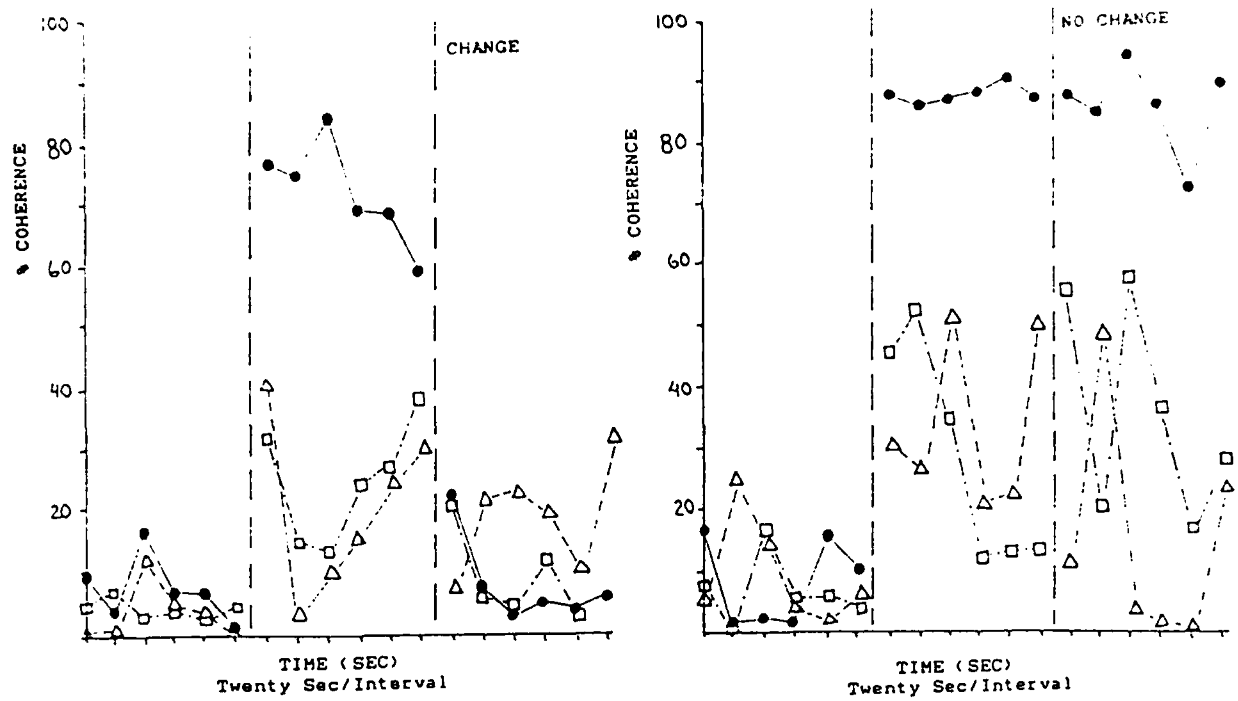


FIGURE 3.
DATA FOR SUBJECT 07



REFERENCES

Kenner, K.M., Junker, A.M., and Levison, W.H., 1985. A Linear Dynamic Model For The Visual-Cortical Evoked Response System. Proc. of NAECON 1985, IEEE Pub. NY, NY. Vol II, pp. 861-867.

Spekreijse, Henk. 1966. Analysis Of EEG Response In Man Evoked By Sine Wave Modulated Light. Dr. W. Junk Publishers. The Hague, Netherlands. 152 pp.

Wilson G.F., and O'Donnell, R.D. 1981. Human Sensitivity to High Frequency Sine Wave and Pulsed Light Stimulation as Measured by the Steady State Cortical Evoked Response. Tech. Report AFAMRL-TR-80-133.

Rescinding Vocational Choice:
A study of attrition of women Coast Guard cadets

Gwendolyn Stevens, Ph.D.
United States Coast Guard Academy
Alyce Hemstreet. B.A.
Eastern Connecticut State University

Abstract

To assess whether or not the personality characteristics of women who become officers differ from those who resign from the Coast Guard Academy a discriminant analysis was carried out for 127 women using data from the California Psychological Inventory, the Edwards Personal Preference Scale, and the Strong Campbell Interest Inventory. A survey forwarded to women who resign also provided information on their continuing career orientation. Data reveal that a cluster of 15 sub-scales adequately discriminates between the two groups at $p < .0001$.

Introduction

Women account for 9.43% of the military personnel in the United States and, while still a minority, their numbers will increase in the future and they will, therefore, be the largest group of women in a non-traditional career path. Although women have always been part of the military, they have never received as much attention, in terms of empirical research, as women in other occupations. Even those other occupations received little attention until the early seventies when women started, en masse, to enter non-traditional positions and an attempt was made to understand what "type" of women wants a non-traditional career (Lemkau, 1979).

Theories dealing with vocational choice are male oriented (e.g., Holland, 1966) and traditional research has analyzed personality characteristics to predict males' vocational choices (Holland, 1966; Osipow, 1968; Roe, 1964). While there is some disagreement between researchers as to whether personality characteristics are sufficient as an explanation for women's vocational choice (Elton & Rose, 1971; Fox & Renas 1977), they do seem to a reasonable place to begin (Chusmir, 1983; Lemkau, 1979, 1983; Bachtold, 1976).

Previous work includes analysis of women Ph.D.s and non-degree-holding workers in traditional "masculine fields," politicians, artists, and ministers. Ph.D. women in "masculine" fields were noted to be more emotionally stable, dominant and resourceful than male Ph.D.s and non-traditional women (O'Leary & Braun, 1972; Helson, 1971). While Patrick (1973) noted that traditional and non-traditional women were similar on the EPPS scales of need for achievement, for autonomy, and for deference, non-traditional women's TAT stories revealed a greater desire for competence and independence. Aloofness has also repeatedly

been noted to be a prime discriminator between traditional and non-traditional women (O'Leary and Braun, 1972; Helson; 1971; Lemkau, 1979). In general non-traditional women have been found to be more dominant, stable, and resourceful than other women (Lemkau, 1979;1983).

Now that the military is integrated, with women and men working, competing, and living together, research on women in the military is likely to become a more popular topic. Accordingly, this project is a first attempt to understand the personality characteristics of women who choose to become military officers (9.39% of the officer corps are women).

Method

Subjects for this study were all the women who entered the Coast Guard Academy as cadets between 1976 and 1979 (the classes of 1980 to 1983).

Procedure: Two procedures were employed for this study.

(1) Discriminant Analysis. Data from the California Personality Inventory (CP), the Edwards Personal Preference Scale (EP), and the Strong Campbell Interest Inventory (ST), were collected for 125 (82.78% of the total number of women had complete data sets) of the women who made up the first four gender integrated classes at the Academy. A discriminant analysis technique was carried out to determine whether there is a constellation of personality characteristics which would separate the group of women who stay and become officers from the group of women who resign.

(2) Survey of resignees. Data for the survey component was solicited by forwarding a questionnaire to all the women who had resigned from the Academy during the classes of 1980 to 1983. The survey instrument used an open-ended format and contained questions which assessed women's inilital reasons for coming to the Academy, reasons for leaving the Academy, and current and future plans, both educational and general. The purpose of the survey was to ascertain if women who resign continue with non-traditional careers. The results are based on a response rate of 70.73%.

Results and Discussion

The discriminant analysis revealed that there appears to be a personality cluster, a combination of 15 subscales from the three personality/interest instruments, which identifies the attriters from the non-attriters with over all accuracy of 76.98%. Correct classification for those who stay and become officers was 80.00% and for those who resigned was 74.20%. As there were no previous indicators, empirical or theoretical, all of the 65 subscale scores were employed in the step-wise analysis, using the SPSS-X process where variables are chosen to minimize the overall Wilks' lambda.

Table 1 displays the Standardized Canonical Discriminant Function coefficients in order of magnitude, and Table 2 displays the group mean responses for the subscales; although several of the group means appear to be quite similar, for the sub scales of domestic arts, endurance, and dominance

there is approximately one-half a standard deviation difference.

Table 2 also contains the results of ANOVA procedures employed to ascertain whether or not there were differences between the groups for the various sub-scales, irrespective of the discriminant funtion analysis. As can be observed, seven of the sub-scales which are on the discriminant function were statistically different for the two groups and an additional two sub-scales indicated statistical differences beyond chance.

The top for variables in the cluster reveal that the women who go on to become officers are most likely to be those who need to be different (the negative relationship for the achievement via conformance score), are usually free from self doubt (the positive relationship for the sense of well-being score), are interested in being perceived in a good light and are concerned about how other react to them (the positive relationship for the good impression score), and need to be a member of a "family" unit where not only help is given but the environment is one of affection and empathy (positive relationship for the succorance score).

TABLE 1

Standarized Canonical Discriminant Function Coefficients

Scale Number and Discriptor	Discriminat Function
CP 13 Achievement via conformance	-.87450
CP 6 Well being	.55414
CP 11 Good Impression	.48649
EP 8 Succorance	.46906
ST 22 Domestic Arts	-.43559
ST 27 Sales	-.43291
CP 17 Flexibility	-.42994
EP 13 Endurance	.42046
ST 19 Teaching	.38555
ST 10 Military Activities	.35811
EP 9 Dominance	.34371
EP 7 Introception	.31246
ST 16 Music	.30321
ST 2 Investigative	-.28235
EP 11 Nuturance	-.29119

The survey data reveal that while more than half of each class resigned from the Academy because they had decided that they did not want to pursue a military career, most of the women are still pursuing an academic degree and have plans to become career women, five have pursued other types of military activities, and only 4 noted that they want a "traditional" life style. The academic programs with which the majority of women are involved are business,

engineering, science, and computer science. The results indicat that although women leave the academy, they do not necessarily change to a traditional career goal even though they choose a more traditional way to accomplish their goal.

Table 2
Group means for the discriminant function subscales

Subscale	Not Resigned N = 60 $\bar{X}$	Resigned N = 66 $\bar{X}$
CP 13 Achievement via conformance	47.32	47.39
CP 6 Well being	44.33	41.37
CP 11 Good impression	45.62	42.64
EP 8 Succorance *	49.53	46.07
ST 22 Domestic Arts *	51.15	56.10
ST 27 Sales *	44.48	47.21
CP 17 Flexibility *	48.42	52.06
EP 13 Endurance *	54.22	47.22
ST 19 Teaching	47.93	47.81
ST 10 Military Activities *	68.48	65.12
EP 9 Dominance *	53.98	48.76
EP 7 Introception	47/20	46.97
ST 16 Music	54.03	53.67
ST 2 Investigative	55.40	54.57
EP 11 Nuturance	49.42	49.49
ANOVA results for sub-scales		
EP 1 Achievement	50.33	55.75
ST 20 Social services	52.18	48.72

(* = p < .05)

While the women who stay and graduate as officers describe themselves as conservative and "not feminists," the results seem to suggest that they are less conformist and more self-directed and independent than are those who leave, even though those who leave also report wanting a career. The results suggest that developing a profile for the "ideal type" women officer is possible. Future work needs to include a similar analysis of males who resign and those who graduate; it is likely that these two groups may be more similar than conventional thinking may suggest.

References

Bachtold, L. M. "Personality characteristics of women of distinction."

Psychology of Women Quarterly, 1976, 1, 70-78.

Chusmir, L. H. "Characteristics and predictive dimensions of women who make non-traditional vocational choices." Personal and Guidance Journal, 1983, 63 43-47.

Elton, C. F. & Rose, H. A. "Sex and occupational choice." Journal of Counseling Psychology, 1971, 18, 456-461.

Fox, H. W. & Renas, S. R. "Stereotypes of women in the media and their impact on women's careers." Human Resource Management, 1977, 16, 28-31.

Helson, R. "Women mathematicians and the creative personality," Journal of Consulting and Clinical Psychology, 1971, 36, 210-220.

Holland, J. L. The psychology of vocational choice. Waltham, MA,: Blaisdell, 1966.

Lemkau, J. P. "Personality and background characteristics of women in male-dominanted occupations: A review." Psychology of Women Quarterly, 1979, 4 221-240

Lemkau, J. P. "Women in male-dominated professions: distinguishing personality and background characteristics." Psychology of Women Quarterly, 1983, 8, 144-165.

O'Leary V. E., & Braun, J. S. "Antecedents and correlates of academic careerism in women." Proceedings of the 80th. Annual Convention of the American Psychological Association,1972, 7, 277 (Summary).

Osipow, S. H. Theories of career development. New York: Appleton-Century-Crofts, 1968.

Patrick, T. "Personality and family background characteristics of women who enter male dominated professions. Unpublished Ph.D. dissertation. Columbia University, 1973. Vol 3415A, p. 2396. Order No. 73-24076.

Roe. A. "Personality structure and occupational behavior." In H. Borow (Ed.) Man in a world of work. Boston: Houghton Mifflin, 1964.

Some Initial Evidence that Catastrophe Theory May Help Explain Cadet Attrition

Michael J. O'Connell*
Steven J. Beatty
Air Force Institute of Technology

Abstract

Catastrophe theory may provide new insights into how and why certain cadets leave the Air Force Academy. Four measures of commitment from June to November over the first year showed no significant differences between those who stayed and those who left at the end of the first year. However, in April, about one month before departure, cadets who departed between April and June had significantly lower commitment than those who stayed.

Background

Cadet attrition at the Air Force Academy has been studied extensively without major advances in understanding and predicting its occurrence. More complicated and sophisticated theory and research is warranted to unravel this complex phenomenon. Catastrophe theory (Zeeman, 1976) is one promising development that may prove useful in advancing our understanding of attrition. This research reports the preliminary analysis of the relationship between commitment and cadet attrition and how catastrophe theory models may be used to explain this relationship.

Catastrophe Theory and Turnover

Catastrophe theory seeks to explain abrupt changes in the behavior of systems whether that be the attack of an angry animal, the crash of the stock market, eating pattern of a person suffering bulimia, or the departure of an employee from an organization (Zeeman, 1976). Sheridan (1985, also see Sheridan and Abelson, 1983) reports two studies of organizational turnover in which a non-linear mathematical model derived from topology is used to explain how commitment and job tension in one case and job tension and group cohesion in the other are related to voluntary termination. As one or both variables change a person is assumed to move along a curvilinear behavior surface characterized by a fold or cusp. For certain combinations of the independent variables the person enters the fold region of the surface where a very small change in, say, commitment or cohesion leads to a large change in behavior, namely, termination.

*The views expressed are the authors' and do not necessarily reflect those of the U.S. Air Force or the Defense Department.

In this research we report how small changes in cadet commitment are associated with cadet's voluntarily leaving the Air Force Academy a short time later. While the data are only suggestive at this point, the analysis does indicate that more sophisticated models may be needed to understand this complex phenomenon.

Methodology

The subjects for this study were 1494 cadets who entered the United States Air Force Academy in 1982. The performance and attrition of this class has been tracked since entry. In addition, surveys were used to collect attitudinal data at various times during their first year when most cadet attrition occurs. A pre-admission survey was mailed in early June 1982 to 705 prospective cadets who had accepted an appointment. Post-admission survey data was collected in early July shortly after entry, in August after basic cadet training, in November 1982, in April 1983 and finally in August 1983. Various subsets of cadets took each survey to preclude undue demands on cadet time. As a result there were repeated measures on various items in the surveys for only a portion of the class.

The attitude instrument used to measure commitment was a 12-item abbreviate version of Mowday, Steers and Porter's (1979) Organizational Commitment Questionnaire. This instrument focused on various components of commitment to the Academy, and each of the 12 items were phased in terms of a statement to which the respondent was asked to rate his or her agreement on the scale of one (strongly disagree) to seven (strongly agree). The overall commitment score for each cadet was computed by averaging across the 12 items.

Administration of this instrument over the first five time periods noted above allowed us to study how changes in commitment relate to subsequent attrition.

Results

To test the overall relationship of commitment to attrition, commitment scores for 323 cadets who completed the June 1982 and the April 1983 surveys were compared for 283 stayers and 40 cadets who voluntarily departed after April 1983. The results of this analysis are shown in Table 1. Both stayers and leavers showed a significant decline in commitment over the 10 month period though the decline was larger for those who left. Those who departed after April had a significantly lower commitment in April than those who stayed. Previous research shows that that those who stay with an organization generally show constant or increasing commitment over time. The large decrease in commitment for stayers is probably due to the extremely high levels of commitment all cadets had upon entering and the sobering effects of the first year's rigorous training.

Table 1
Changes in Commitment Over Time

Group		June 82	April 83	t (Jun vs Apr)
Stayers	mean	6.239	4.570	2.97
(n = 283)	std. dev.	.527	.92	
Voluntary Leavers	mean	6.235	4.156	3.03
(n = 40)	std. dev.	.598	.96	
t (stayers vs leavers)		.00	2.25	

While commitment of both stayers and leavers decreased during the first 10 months, the comparisons in Table 1 do not show when the divergence of the two groups occurred. A further analysis was conducted to track commitment for both stayers and leavers at three time periods between June and April. Table 2 shows average commitment scores for the two groups. In this comparison stayers include all who completed a given survey and who were still enrolled as of July 1983; leavers include those who completed the indicated survey and left between April and June 1983.

Table 2

Average Commitment Scores During the First Year at Academy

Survey	Stayers[1] (as of July 83)	Voluntary Leavers (April to June 83)	Significance of Difference
June 82	6.239 (n = 338)	6.235 (n = 43)	n.s.
July 82	5.625 (n = 429)	5.561 (n = 40)	n.s.
August 82	5.819 (n = 513)	5.880 (n = 40)	n.s.
November 82	5.078 (n = 419	5.014 (n = 42)	n.s.
April 83	4.570 (n = 283)	4.16 (n = 27)	$p<.05$

[1] Since different subpopulations of cadets took the commitment instrument at the different administrations, there are different size groups of stayers and leavers.

Discussion

Between November and April a decline in commitment was experienced by both stayers and leavers, but the larger decline for some apparently pushed them to a definite decision to leave the Academy. They had reached a point where a more radical or abrupt response was their appropriate coping behavior. As Sheridan (1985) explains, other, less radical behaviors such as poor performance and absenteeism are typical responses short of termination.

Many other factors were operating on the decision to leave the Academy. We have focused on only commitment, but realize there are many aspects of the Academy that lead to high or low cadet commitment. These other factors may explain why others, though experiencing low commitment also, did not leave. Each cadet probably has her or his own point beyond which termination is warrented. How individual differences effect a person's response to small changes in commitment needs further research (Sheridan and Abelson, 1983).

These initial results are only suggestive at this time. Catastrophe theory warrents further study in which specific mathematical models incorporating several variables are used to develop behavior response surfaces for groups of cadets. Such models will advance understanding of both the attrition process over time and the interaction of key factors that influence it.

References

Mowday, R.T., Steers, R.H., & Porter, L.W. The measurement of organizational commitment. Journal of Applied Psychology, 1979, 14, 224-227.

Sheridan, J.E. A catastrophe model of employee withdrawal leading to low job performance, high absenteeism, and job turnover during the first year of employment. Academy of Management Journal, 1985, 28, 88-109.

Sheridan, J.E. & Abelson, M.A. Cusp Catastrophe Model of employee turnover. Academy of Management Journal, 1983, 26, 418-436.

Zeeman, E.C. Catastrophe theory. Scientific American, 1976, 234, 56-83.

A STUDY OF ENLISTED ATTRITION IN THE U. S. COAST GUARD: 1983-1984

Stephen B. Wehrenberg
Office of Personnel, USCG
November, 1985

ABSTRACT [1]

Reports the results of a survey of 364 people whose enlistments expired between April and October 1984; stated intent was tested against actual "stayed-left" behavior. Results were coupled with performance marks, adding the much needed dimension of performance to attrition research. The constructs of satisfaction, commitment, and investment were analyzed over time. The most important issues were assignment and lack of control over professional development (career paths). Compensation issues were less important but investment in post-service benefits was cited as a major reason for staying. In general, high performers are dissatisfied with the degree of influence they have over their own futures.

INTRODUCTION

In 1979 the Coast Guard Office of Personnel commissioned a study of enlisted attrition. That study (Wehrenberg and Patterson, 1981) was somewhat unusual in that it did not focus on compensation issues. The results were not surprising from a theoretical perspective, but did startle some Coast Guard planners. For example, it was shown that people who received high reenlistment bonuses "liked" the Coast Guard less than those who received lower bonuses, and less than they had before they received the bonus. Selective reenlistment bonuses were keeping people, but was it for the "right" reasons? Were they the people who would deliver the most productivity?

Three classes of factors were identified from that study as driving the decision to stay or leave. These three factors formed the basis of the current research:

Satisfaction -- a global attitude toward the Coast Guard; the weighted sum of the feelings about individual experiences while in the Coast Guard, including elements of job and non-job satisfaction.

Investment -- the notion that an individual has "too much invested" in the Coast Guard to be able to leave; this is compensation related. Either the individual has too much invested in retirement to leave, or an individual's skills are so specific to the Coast Guard or to the service that it would be hard to find a market for those skills outside the Coast Guard. Investment, at some point, becomes a feeling that "I can't leave ..." no matter how much he or she would like to. A high sense of investment coupled with low satisfaction leads to a phenomenon we describe as spiritual turnover, in which the individual is still with us, but in terms of attitude and commitment has "departed."

Commitment -- a sense of identification with the goals and missions of the Coast Guard, and with the Coast Guard itself. A person who exhibits high commitment is able to connect nearly any task, no matter how trivial, to the accomplishment of unit and Coast Guard goals.

----

1. Due to the limits placed on this paper, references to statistical processes will be omitted. Depending on the type of test, a significance level of either $p<0.05$ or $p<0.01$ was used. Any real differences reported between groups have met this test. Citations in the text are available from the author.

Following from Dubin's theory of "central life interest," these attitudes appear to be mediated by that aspect of an individual's life that is most important; for example, if family and geographic stability are important to an individual, assignment policies become important and the source of either global satisfaction or dissatisfaction. If security is important, anything that threatens that security becomes a dissatisfier (e.g. proposals to reduce the retirement benefit). If the most important thing in a persons life is identified as the Coast Guard itself, any changes to the status quo are threatening and create dissatisfaction. We have termed this moderator variable "organizational life priority" and consider it to be a filter between experience and attitude.

It was felt by the researchers and by planners in the Office of Personnel that the best mix for the Coast Guard would be to have people who were committed to the Coast Guard as an organization, who were reasonably satisfied with their experiences and positions, and had just enough sense of investment to "carry them through" the bad days. Unfortunately, such was not the case. Investment accounted for over 60 percent of the variance when comparing those who stayed with those who left. It seems that we were doing everything possible in personnel policy to instill a sense of investment in people.

The purpose of the current research is twofold. First, since voluntary attrition is at an all time low, it would be valuable to reexamine the attitudes that were captured in 1980. This would give us a sense for the variability of these attitudes across differing economic conditions. The second, and even more promising, element is that the data capture of this study coincided with the first round of appraisals under the new Enlisted Performance Evaluation System (EPES). By capturing both attitude and performance data we can examine the workforce in a two dimensional structure, retention and performance, giving us the ability to compare attitudes and to develop policies that target those high performers who plan to leave.

An expanded version of the survey used for the 1980 study was used in this research. A section on compensation issues (based on the work of the Fifth Quadrennial Review of Military Compensation) was added, including benefits and other forms of payment in kind.

RESULTS

As was noted in the 1980 study, average "Commitment to the Coast Guard" is in the mid-range of the possible responses, increases somewhat with time, and drops as the sample approaches the first retirement opportunity.

Satisfaction does vary with time, by rating, by other demographic variables, and by the "object." A global measure of satisfaction shows that satisfaction starts high, bottoms at the four year point, and gradually increases out to about 12 years, where it begins to drop slightly to beyond 26 years.

Global satisfaction is not the best indicator of reenlistment intentions. Two individual satisfaction scales tend to be better predictors, Satisfaction With the Job, and Satisfaction with Future Assignments. Satisfaction with the job includes elements of altruism (contribution to lifesaving, etc) as well as the conditions surrounding the job (right tools, etc.). Satisfaction with future assignments includes knowing what options are available as well as the assignment process itself.

The greatest variance across time is seen in the investment scale. The average response starts low (little invested) and increases dramatically to about the 12 year point, where is levels off until about the 20 year point. At the 11-12 year point the lure of retirement benefits that would be lost by leaving is strong. Up until that point bonuses and security drive investment. As was the case in the 1980 survey, investment accounts for much of the variance in the reenlistment decision. Many people indicated that they would like to leave if only they didn't stand to lose so much.

Canonical discriminant analyses were used to differentiate the four performance/continuance

cells, and the results of these analyses have been useful in examining the overall structure of attitudes among the four cells for each time in service group. However, for the purposes of this paper, I will focus on the differences between those who stay and those who leave among the high performers. If we can determine what issues differentiate leavers from stayers among high performers, and the issues are of such a nature that adjustments to policy can target the leavers without jeopardizing the attitudes and intent of those who already plan to stay, we may be able to keep more high performers than at present.

UNDER FOUR YEARS IN SERVICE

The OLP scales indicate that those who stay report that the Coast Guard is important to them. Specialty is more important to those who left. This could point out a dichotomy between those who consider themselves generalists and those who consider themselves specialists. Confirming this, those who stayed report higher commitment to both the Coast Guard as an organization and to the goals and missions of the Coast Guard.

Those who left reported dissatisfaction with assignment policies in general, dissatisfaction with their potential future assignments, and would like more control over the assignment process. Those who stayed reported greater satisfaction with the geographic area to which they were assigned.

High performers who left also reported lower satisfaction with the equipment and tools they used to do their job, and that the "way things turned out" for them was different from what they expected when they entered the Coast Guard.

Those who stayed reported a greater sense of investment than those who left. From interview data, investment for the under four years in service group was closely tied to the perception of marketability. Their investment was stated in terms that they would "lose ground" if they got out at the end of their enlistments. Other studies point out that many of those between three and four years in service have already extended their enlistments by three years or more to receive a Selective Reenlistment Bonus (SRB) and are therefore "invested" financially as well. An additional factor seen from the interview data may be that as a person reaches the end of the first enlistment, he or she is likely to be recently married, the woman of the pair may be pregnant, and the thought of losing medical benefits at just the point where a major expense is expected creates even more investment.

On the compensation variables that differentiated stayers and leavers (among high performers), those who stayed felt that they were losing money in the long run, and that SRB policies were not made in their best interests. Stayers also report that basing pay on the number of dependents one has is a good idea. Those who left were dissatisfied with the medical care they received while in the service. The interview data indicated that it was not the quality of medical care per se, but the way they were treated, the long delays, and the overall impersonal nature of medical care that dissatisfied them. Overall, those who chose to stay were slightly more senior, and more likely to be married and have more dependents.

To contrast these people with low performers, the low performers were much more likely to be concerned with issues of compensation, their perception of marketability, and dissatisfaction with their specialty. The low performers also felt that vesting pay in the billet was a much better compensatory mechanism than vesting pay in the performance of the individual (and no wonder).

BETWEEN FOUR AND EIGHT YEARS IN SERVICE

This group was divided on fewer issues, but the strength of the differences was as great as for the previous group. The OLP scales that best differentiated high performing stayers and leavers were OLP Coast Guard and OLP Leadership. Those who stayed indicated (like the under-four group) that their most important organizational priority was the Coast Guard itself. Those who left indicated that the Quality of Leadership was more important.

This explains the finding that those who stayed felt more satisfied with the treatment they had receive by those senior to them, and the supervision and the supervisory support they had experienced. Those who left indicated dissatisfaction with these issues.

The other issues that differentiated stayers from leavers were satisfaction with control over the job, satisfaction with future assignments, and satisfaction with the tools and materials needed to do the job. In all three cases, those who left were less satisfied.

Differences between stayers and leavers among these high performers could be found on only one compensation issue: the perceived adequacy of retired pay. Those who stayed felt that retired pay was not adequate; those who left reported retired pay as adequate, but not relevant. This indicates that the issue of post-service benefits becomes important to people somewhat earlier than had been reported in other studies. Those who stayed were slightly older, slightly more senior, likely to be married and have more dependents, and slightly more educated than those who left.

Comparing these high performers to the low performers in this group, the low performers were much more concerned with compensation issues (base pay, SRB, sea pay, dependent allowances, and pro pay), medical care for themselves and their dependents, and the low performers who stayed scored high on the investment scale.

OVER EIGHT YEARS IN SERVICE

The issues of greatest importance to those over eight were OLP Geographic Area and, again, OLP Coast Guard. Those who stayed stated that the Coast Guard was important to them; those high performers who left were more concerned about their geographic location. This is again supported by responses to the scales concerning assignment policy.

Those who left were less satisfied with assignment policies in general, less satisfied with their potential for future assignment, and less satisfied with the amount of control they have in the assignment process. The leavers also reported low satisfaction with the amount of control they have over their jobs, and report a high "sense of helplessness."

There were more compensation issues evident in this group. In each case, those who stayed were more satisfied with SRB policies, VHA, travel allowances, medical care for themselves and their dependents; felt that CHAMPUS had worked for them; and were more satisfied with housing. The people in this group who stayed in the Coast Guard tend to be slightly younger (less time in service), slightly more senior, more likely to be married, and more educated than those who left.

Comparing this group to low performers with over eight years in the service, we find that the same issues differentiate those who stayed and those who left. The low performers were differentiated on the issue of satisfaction with the job itself, those low performers who stayed being more satisfied. The finding that the same issues differentiate reenlistment decisions for both high and low performers supports the thesis that as people "age" in the service, or as the population becomes more homogeneous, their attitudes become more and more alike.

DISCUSSION AND CONCLUSION

In contrast to the 1980 study, in which compensation issues played a role, the single greatest issue arising in this research is assignment. Considering that "control over the job" came up in each group, this may be generalized to indicate that, at least among high performers, the perception of control is an important issue.

This finding is supported by past research conducted by the Navy. In one survey (CNO ESQ83), the primary reasons people gave for leaving were that they disliked being separated from their

Implications of the Results of Recent Turnover Research
For Air Force Policy

Thomas W. Watson
Air Force Human Resources Laboratory
and
Victor H. Appel
The University of Texas at Austin

Abstract

Recent research under the auspices of the Air Force Human Resources Laboratory (AFHRL) has identified factors influencing turnover intent and behavior of first- and second-term enlisted personnel in eight occupational specialties. Implications for Air Force policy of the results obtained are the topic of this paper. The results are striking and have clear policy implications. Relatively few factors consistently accounted for up to 68% of the variance in turnover and related outcomes, an accomplishment unparalleled in the literature. Results generalized across occupations, and were stable upon cross validation. All but one of the factors identified are under management control. Thus, they lend themselves to remedial action. Each factor is discussed from the perspective of what actions management might take to induce valued airmen to remain in the Air Force. As a general rule, enhancing commitment appears to be more critical than increasing compensation as a managerial intervention strategy.

Turnover is a costly outcome for organizations. Thus, employers are often concerned about keeping valued employees in whom they have invested considerable resources. Unfortunately, until recently, too little has been known about the factors influencing turnover decisions, especially those of a perceptual nature. To improve upon this situation, the present investigators developed, empirically validated, and tested an advanced, hybrid model of turnover on a world-wide sample of 3,998 first- and second-term enlisted personnel in eight occupational specialties. Detailed explication of this research is provided by Watson, 1985; Watson, Appel, and Brown, in review; and Watson, Appel, and Black, in review. The present paper focuses upon the practical implications[1] deriving from a model of turnover which emerged, called the Phase III Model.

The Basis for the Implications Offered

The Phase III Model, which serves as the basis for the implications, is represented in Figure 1. In the investigation referenced above, predictors in this model were so strong that together, they typically accounted for between 50% and 68% of the variance in each of four criteria. Further, the results held up remarkably well upon cross-validation. Shrinkage was usually less than .05. In addition, the findings appear to be generalizable for two reasons. First, the results were quite consistent across the eight

---

[1] The views expressed here are those of the authors and do not necessarily reflect the views of the United States Air Force or the Department of Defense.

Air Force specialties (AFSs) examined. Second, demographic variables did not serve as potent moderators, producing differential outcomes. Thus, the results obtained could be represented by a single Phase III Model (Figure 1). The simplicity of this model is dramatic. A relatively small subset of variables from an earlier, more complex model emerged as potent predictors of the first intermediate criterion. Predictors were retained in the Phase III model if they were statistically significant ($p < .05$), and if they contributed at least .02 to the increment in explained variance. Looking at the model in a time-sequenced fashion, moving from left to right, note that progressively fewer variables predict subsequent intermediate or final criteria. Of all the initial predictors in the model, only Availability of Alternatives is not under direct or indirect management control. All of the remaining predictors can be influenced by management and are directly predictive of the first intermediate outcome, Commitment to the Air Force. This construct turned out to be an extremely important criterion variable since it proved to be a direct determinant of both Thoughts of Leaving and Search and Propensity to Stay. Airmen who scored high on Commitment to the Air Force tended not to think about leaving and tended not to engage in search behavior. They also expressed a stronger propensity to stay in the Air Force than did their counterparts who scored low on this criterion. Although Commitment to the Air Force was not directly related to the behavioral Stay/Leave criterion, it was indirectly related to this final outcome through the more proximate criterion, Propensity to Stay. Since Commitment to the Air Force was demonstrated to have such a strong and almost exclusive impact on later outcomes in the stay/leave decision process, the present authors assert that it should be cultivated to induce

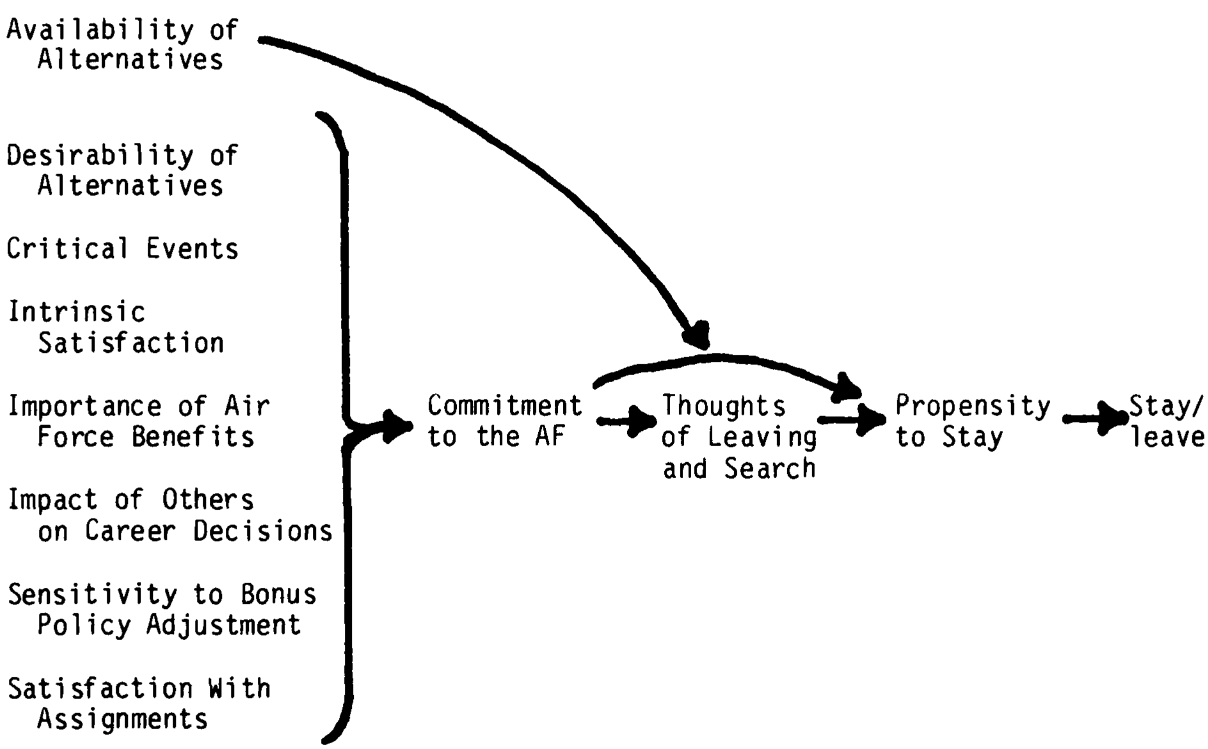


Figure 1. The Phase III Model of turnover upon which implications are based.

valued airmen to remain in the Air Force. Fortunately, the research conducted provides insights into how this might be accomplished. Below, the implication that managers should foster commitment is elaborated upon, followed by a discussion of specific actions management might take to enhance commitment. This discussion focuses upon the predictors on the left side of the model which influenced Commitment to the Air Force.

Implications

From the large data base produced in the investigation upon which this paper is based, a major Air Force policy implication emerged. This implication is as follows: as a managerial strategy, increasing commitment may be as important as maintaining an adequate level of compensation. Although adequate, equitable compensation must be maintained, the data point to the fact that Air Force enlisted personnel are very likely to respond to reenlistment efforts appealing to their sense of commitment to the Air Force as an institution and as a way of life. This is consonant with the argument by Moskos (1976) a decade ago that an instititutional orientation among service members should be developed. Given the results of the current investigation, the issue becomes a question of how might commitment to the Air Force might be enhanced. The emphasis on a commitment-oriented strategy contrasts somewhat with current Air Force practice, despite interest among Air Force managers in Moskos' institutional versus occupational distinction. Certain aspects of compensation obviously influence retention. For instance, 87% and 76% of the respondents studied indicated that pay/allowances and bonuses, respectively, were either very important or somewhat important in influencing their decision to stay or leave. Understanding the importance of compensation, Air Force managers have emphasized factors such as pay and bonuses, the latter fluctuating with cyclical changes in the economy. Unfortunately, if pecuniary factors are given too much emphasis, recruits may be attracted or incumbents may be developed who are more interested in what they can get from the Air Force, than in what they can give to it. Thus, on the basis of their research findings, the current investigators conclude that although adequate compensation must be maintained, this economic emphasis should be augmented by a psychological thrust aimed at cultivating Commitment to the Air Force.

To cultivate commitment, policy makers need to take remedial action with regard to the following predictors: Desirability of Alternatives, Critical Events, Intrinsic Satisfaction, Importance of Air Force Benefits, Impact of Others on Career Decisions, Sensitivity to Bonus Policy Adjustment, and Satisfaction with Assignments. Examples of how each factor can be given increased managerial emphasis are discussed below.

Desirability of Alternatives (DA). It is evident that many enlisted personnel perceive there to be desirable career alternatives to continuing in the Air Force. To counter these perceptions, the Air Force needs to increase its own perceived desirability. To accomplish this, Air Force managers need to become more familiar with personnel policies of the best civilian firms. They could then alter Air Force policies, within sometimes unavoidable constraints, to become more competitive with the best civilian firms. Once the Air Force is made a better place for a career, the assets of such a career will need to be marketed more aggressively to potential recruits and to Air Force members.

Critical Events (CE). Critical events are events which stand out in an airman's recollection as having either a particularly positive or negative

impact on the decision to stay or leave. The critical events examined in this study are listed below in descending order of their influence. Where the direction of influence is not intuitively obvious, it is indicated in parenthesis. Critical events are: (1) getting a promotion; (2) increase in pay; (3) a good performance appraisal; (4) an offer/advertisement from a civilian firm; (5) not getting a promotion; (6) a decrease in pay (7) a poor performance appraisal; (8) a conflict with spouse's job; (9) a change in supervisor (stay); and (10) a change in co-workers (stay). It should be noted that although CE was a strong predictor of Commitment to the Air Force, such events occurred, and were acknowledged to have an impact on stay/leave decisions, of less than half of the participants in the study.

Intrinsic Satisfaction (IS). The fact that IS was an important determinant of Commitment to the Air Force suggests that strategies which promote IS would increase retention and reduce the need to rely so heavily upon extrinsic motivators. Managers at the Manpower and Personnel Center (MPC), and at the Leadership and Management Development Center (LMDC), have already made great strides in this area. One additional action which can be taken is to redesign jobs to increase the intrinsic value of work (see Watson & Zumbro, 1977). Consonant with efforts to redesign jobs, the current national awareness of the characteristics of excellent corporations (see Peters and Waterman, 1982) underscores the need to provide personnel with autonomy and opportunities for innovation, even within highly structured, bureaucratic organizations. Doing so should enhance IS.

Importance of Air Force Benefits (IAFB). Air Force managers have long been aware of the importance of benefits, some of which are unique to the Air Force. They should certainly continue their efforts to prevent their erosion. The benefits identified in this study, in descending order of their importance to the respondents queried, are as follows: (1) medical/dental care; (2) job security; (3) off-duty education; (4) leave policies; (5) bonuses; (6) commissary/exchange privileges; (7) early retirement opportunities; (8) travel opportunities; (9) morale, recreation, and welfare facilities; and (10) commissioning programs for airmen.

Impact of Others on Career Decisions (IOCD). The strong influence of family and friends on stay/leave decisions suggests that organizational socialization begins prior to Air Force entry. It also suggests that Air Force members continue to be influenced by important people in the external environment once they are on the job. Therefore, the Air Force could make special efforts to recruit personnel with Air Force or other military backgrounds since they would be familiar with expectations and would have family/friends who tend to favor military service.

The Air Force has demonstrated a concern for the perceptions and well being of military family members. Action should continue to be taken to monitor and improve spouse satisfaction with the Air Force. Besides monitoring spousal attitudes, the Air Force could increase their efforts to aggressively market the Air Force way of life to family members. To the maximum extent possible, Air Force managers should develop policies which are sensitive to the independent career aspirations of spouses. The situation faced by couples who are both service members, addressed under the join-spouse program, is a special case which needs continued attention. Efforts to co-locate military spouses wherever possible should be supported. In addition, initiatives to further reduce the frequency of PCS moves should be considered.

Sensitivity to Bonus Policy Adjustment (SBPA). The Air Force's policy of providing bonuses in response to supply and demand factors may be backfiring. This strategy, and the cyclical nature of the economy, result in bonuses being given at one reenlistment point and taken away at the next. The results of this study indicate that reelistment bonus reductions due to cyclical economic changes produce frustration and resentment, which increase the desire to leave when opportunity permits. Ideally, bonuses should be given to individuals, based on merit. If bonuses are provided across AFSs in response to factors irrelevant to performance, such as variations in supply and demand, they should at least be as stable as possible.

Satisfaction with Assignments (SA). For many years, much management attention has been devoted to developing and refining an equitable assignment system that can handle the huge number of actions occuring each year. Innovative solutions to assignment system shortcomings should be considered. For instance, a computerized assignment system could be integrated with a sophisticated career management system like the one currently used by IBM and other respected civilian corporations.

Summary

It is fortuitious that the results of the extensive retention study carried out by the investigators and their colleages (Watson 1985; Watson, Appel, & Black, in review; Watson, Appel, & Brown, in review) resulted in outcomes which lend themselves relatively easily to corrective actions. The investigators have stressed the important central finding that Commitment to the Air Force is a particularly important factor which should be cultivated. The investigators have identified factors influencing commitment which indirectly influence decisions to stay or leave. Illustrative suggestions for change have been offered which should stimulate readers to consider these and other strategies to improve the probability of retaining valued airmen.

References

Moskos, C. C. (1976). Armed forces in American society: From institution to occupation. In Proceedings of the 5th Symposium on Psychology in the Air Force. Colorado Springs: USAFA.

Peters, T. J., & Waterman, R. H. Jr. (1982). In search of excellence. New York: Harper and Row.

Watson, T. W. Multivariate models predicting Air Force turnover and retention which emphasize the desirability of alternatives. Unpublished doctoral dissertation, The University of Texas at Austin.

Watson, T. W., Appel, V. H., & Brown, R. H. (in review). Empirical validation of an advanced, hybrid model of turnover. Journal of Vocational Behavior.

Watson, T. W., Appel, V. H., & Black, D. E. (in review). An advanced, hybrid model of turnover: Its fidelity, generalizability, explanatory power, and stability. Journal of Vocational Behavior.

Watson, T. W., & Zumbro, P. A. (1977, October). Job enrichment: Evaluation, with implications for Air Force job redesign (Report No. AFHRL TR-77-56). Brooks AFB, TX: Occupation and Manpower Research Division, Air Force Human Resources Laboratory. (NTIS No. AD-A050 827)

New Directions
A Survey of Separating Officers

2d Lt Kevin A. Ashley, USAF
Charles H. Hamilton
Air Force Military Personnel Center

Abstract

This paper examines information gathered from personnel voluntarily separating from the Air Force. The methodology used in conducting the direct-mail survey is addressed, as well as a discussion of the results, including timing of the separation decision, satisfaction with Air Force experiences, and future plans outside the military.

Introduction

The Air Force (AF) Personnel Survey Branch is responsible for conducting and supervising surveys of Air Force personnel in a continuing effort to provide AF senior leadership with information essential for informed decision making on personnel policies and programs. The Survey Branch accomplishes the majority of its studies using direct-mail surveys, a method in which surveys are distributed from the Survey Branch directly to the participants at their duty address, and then mailed back to the Survey Branch upon completion. Depending on the survey subject, time of year, and number of items included in the survey, response rates may vary from 50% to 85%.

In 1979, in an effort to identify factors affecting voluntary separations, a survey program was developed and implemented Air Force-wide for use in improving retention of Air Force personnel. This program, established in the Consolidated Base Personnel Office (CBPO), affords all voluntarily separating members the opportunity to complete the "exit" survey as they complete their outprocessing paperwork. Responses are then packaged monthly and forwarded from the separation units within the 124 CBPOs AF-wide to the Survey Branch for analysis. Since its adoption, this program has provided very useful and important data, but total survey returns have been received from only 8%-15% of the total group of voluntary separatees.

Having traditionally enjoyed better return rates with direct-mail surveys, a decision was made to test a new survey of separating personnel using the direct-mail method. While the traditional survey would continue to be used, the direct-mail survey would be employed to enhance these data by providing additional information on other separation related issues.

Method

One of the inherent drawbacks of employing this centralized approach to conducting a survey of separating personnel is the

large number of annual voluntary separations. In 1985, 2,615 officers and 36,725 enlisted personnel voluntarily separated from the Air Force. To simplify the test of this new survey, only officers with a verifiable date of separation in their personnel records were included in the study. The target population was established as officers with dates of separation between 1 May and 31 October 1985. This dramatically decreased the size of the potential subject pool to 720 individuals.

The survey was composed of seventy-four multiple choice items and one open-ended item, and addressed the timing of the separation decision, with whom the separation decision was discussed, and members' satisfaction with a variety of job factors during their tenure in the Air Force.

To mirror the traditional survey where possible, all officers who would be separating between May and October were given the opportunity to participate in the survey, and were thus forwarded a survey package early in April 1985.

Results

Five weeks after mailing, responses from 55% of the target population had been received. Considering this large number of returns, responses were representative of the target population at the 95% confidence interval, $\pm$ 5% error.

The total group of returns was composed entirely of captains and majors, of whom 79% were males. 28% of the respondents had served less than six years in the military at the time of their separation, 45% between six and nine years, and 27% more than nine years. Concerning their source of commission, the plurality of respondents (44%) received their commissions through the Reserve Officer Training Corps (ROTC). 29% received commissions from Officer Training School (OTS), 24% from the United States Air Force Academy, and 3% through other commissioning sources. The group was composed of 59% support officers, 32% pilots, and 9% navigators.

Concerning the timing of the separation decision, about one-fourth (26%) made their decision to separate from the Air Force more than a year before submitting their application papers. Another 17% made their decision 7-12 months before submission; 39%, 1-6 months before; and 18% within a month of applying for separation.

When asked with whom they had discussed their decision to separate before submitting their application, 80% had discussed their decision with friends and/or co-workers, 69% had discussed it with their supervisor, and 50% had discussed their decision with their commander.

Using a seven-point scale ranging from very dissatisfied to very satisifed, respondents were asked to indicate their level

of satisfaction with twenty job-related factors during their stay in the Air Force. Overall, respondents had been the most satisfied with the quality of their co-workers, pay, and job responsibility while in the Air Force. On the other hand, they had been the least satisifed with geographic stability, quality of leadership and supervision, and opportunities to exercise creativity.

Again using this same set of job factors, the survey asked participants to classify each item according to whether they thought it could be better met in the Air Force or in the civilian sector. Overall, most job factors were considered to be potentially better met in civilian employment, especially geographic stability, "say" in base/location of assignment, and "say" in particular job assignment.

Several questions addressed the participants future plans once they separated from the Air Force. Slightly more than half, though, were not leaving the military entirely as 54% indicated they were planning to join the national guard or the reserves. Concerning future plans outside the military, the majority (79%) said they were planning to work, half of these already having a job arranged, mostly with private industry.

The open-ended item asked participants to indicate the main reason they were separating from the Air Force. Although items requiring "write-in" responses are often skipped by survey participants, in this case, more than three-fourths (78%) of the respondents responded to this item and included additional written comments. These comments generally fell into four broad categories: changes in career goals, family factors, leadership and supervision in the Air Force, and economic factors. Half (51%) of the respondents indicated they wanted to pursue career goals not available to them in the Air Force (i.e., start their own business, go to school, etc.). Also, 42% commented that their decision was influenced in part by what they perceived as poor leadership and an inflexible assignment system. Similarly, 37% indicated that they wanted to remain in one geographic area, and/or have a more stable family environment. Finally, 19% said their decision to separate from the Air Force was due to the higher salary they could receive for comparable work in the private sector.

Discussion

The direct-mail approach did in fact produce a significantly higher response rate. Although separating personnel tend to be "looking ahead," and historically have not taken the time to provide information regarding their separation decision, these respondents completed the direct-mail questionnaire and many provided detailed, candid write-in comments.

Overall, the survey results seem to indicate that these separating officers had enjoyed their years in the Air Force, but felt the civilian sector offered several distinct benefits,

especially in the areas of pay, geographic stability, and "say" in their jobs. Still, many of the respondents were not making a complete break with the military way of life as over half were planning to join the national guard or reserves. This statistic may be influenced by the fact that the military offers certain opportunites that are not available in the civilian sector, such as flying a jet fighter, as well as the opportunity to serve one's country.

THE 1984 ACTIVE DUTY PSYCHOLOGISTS SURVEY: RETENTION IMPLICATIONS

A. David Mangelsdorff, Ph.D., M.P.H.
Division 19 (Division of Military Psychology)
American Psychological Association

ABSTRACT

What factors affect the decision to remain in the military or to leave the service? The Division of Military Psychology (Division 19) of the American Psychological Association conducted a survey of all military psychologists on active duty in 1984. The purpose was to document issues of concern to military psychologists. Issues were addressed from two perspectives: long term motivators and from the degree of satisfaction felt. Regression equations were developed to predict the response to a 7-point criterion item "Likelihood remain until eligible for retirement." Separate sections were developed to describe the findings from the military groups. The issues of most concern in 1984 dealt with the establishment of a professional career and a professional identity. There was a significant relationship between likelihood remain until eligible to retire and years of active military service completed (more years completed: greater likelihood remain to retire). Prior military service before becoming a psychologist was also important. Differences between clinicians and research psychologists were noted.

INTRODUCTION

The factors that affect the decision to remain or to leave the military are numerous. The issues have been addressed at varying times (Hedlund, 1968; Jacoby, 1970; Mangelsdorff, 1978, 1984; Maynard, 1982; Meling, 1982; Mitchell, Keeth, and Wiekhorst, 1982; Mitchell and Keeth, 1984; Murray, 1978; Rosenbach and Gregory, 1981). The recurrent issues identified as influencing military health care professionals to leave the service were: inadequate pay, possibility of nondirect patient care assignments, lack of participation in decisions affecting their careers, lack of sense of belonging to the military community, inadequate integration into military social life, poor equipment and support facilities. The current study was part of the 1984 survey of active duty military psychologists conducted by the Division 19 of the American Psychological Association. The purpose was to document issues of concern to military psychologists.

METHOD

The Division 19 survey instrument was mailed to all military psychologists on active duty in 1984. Respondents were asked for demographic data, attitudes toward a military career, and attitudes toward a variety of issues. The issues were assessed from two perspectives: long term motivator factors and from the degree of satisfaction felt. Attitude statements were assessed using 7-point Likert scales (1 = minimum to 7 = maximum).

A regression equation was developed to predict the response to the 7-point criterion item "Likelihood remain until eligible for retirement" (scale endpoints were 1 = low probability to 7 = high probability). From the pool of demographic section responses and Long Term Motivators, items having content specifically dealing with military career, psychology, and/or military identity were identified. The items selected in the stepwise linear regression analyses best reflect issues affecting the decision to remain in the service.

RESULTS

Overview.

Issues of concern and issues affecting retention until eligible to retire from the service will be presented. A composite "researcher" profile, a composite "clinician" profile, and an overall military psychologist will be discussed.

COMPOSITE MILITARY CLINICIAN PROFILE.

The composite clinician is a male, O3 rank, aged 36.2 years, with 8.3 years military service completed, 2.5 years prior military service before becoming a psychologist, and in the present assignment for 19.3 months. He has a doctorate conferred (96% of clinical respondents did). He ranks himself highest as a psychologist. He reports being highly committed to his work.

Regression Equation Development.

The regression equation developed to predict likelihood remain until eligible to retire was significant: $F=58.46$, $df = 7/242$, $p = .00001$; multiple $r = .792$; $R^2 = .628$). The most salient issues for the composite clinician in 1984 (with betas) were:

"Total years active military service completed"	(.58)
"Commitment to the military"	(.32)
"My career progression"	(.10)
"Extent work in field of psychology in which were trained"	(.14)
"Likelihood anticipate being promoted"	(.09)
"Having retirement benefits available"	(.10)
"Months in present assignment"	(-.08)

THE COMPOSITE MILITARY RESEARCHER.

The composite researcher is a male, O3 rank, aged 36.1 years, with 9.9 years military service completed, 1.8 years prior military service completed before becoming a psychologist, and in the present assignment for 24.8 months. He has a 70% probability of having a doctorate conferred. He ranks himself highest as a military officer. He is highly committed to his work and is most concerned about the availability of retirement benefits as a retention factor.

Regression Equation Development.

The regression equation developed to predict likelihood remain until eligible to retire was significant: $F = 16.24$, $df = 10/85$, $p = .00016$; multiple $r = .810$; $R^2 = .656$). The most salient issues for the composite researcher in 1984 (with betas) were:

"Total years active military service completed"	(.41)
"My professional identity as a military psychologist"	(.30)
"Extent feel a sense of membership in the service"	(.22)
"The availability of incentive pay"	(.18)
"Extent being a psychologist offers opportunity for advancement/promotion in the military"	(.18)
"Extent responsible for own career development"	(-.20)
"Having retirement benefits available"	(.19)
"Desirability of an equivalent amount of time for advanced professional training as a substitute for suggested/required military training"	(-.23)
"Years prior active military service before becoming a psychologist"	(.17;)
"Having opportunity to receive post-doctoral training while on active duty"	(.15)

THE COMPOSITE MILITARY PSYCHOLOGY OFFICER.

Regression Equation Development.

The regression equation developed to predict the likelihood that the composite military psychology officer would remain until eligible to retire was significant: $F = 70.79$, $df = 7/345$, $p = .00001$; multiple $r = .767$, $R^2 = .589$). The most salient issues for the composite military psychology officer in 1984 (with betas) were:

"Total years active military service completed"	(.51)
"Commitment to the military"	(.22)
"Having retirement benefits available"	(.13)
"Likelihood anticipate being promoted"	(.12)
"My personal identity as a military psychologist"	(.11)
"Opportunity to attend resident advance course"	(.09)
"Availability of psychology positions in reserves"	(-.07)

DISCUSSION

The issues of most concern to the active duty psychologists in 1984 dealt with the establishment of a professional career and a professional identity. Though there were variations between the services and specialties, the common themes focused on career development and progression, responsibility, the opportunities for self improvement, opportunities for independent thought and action, availability of retirement benefits, and professional experience. The promotion criteria, professional military education, and elements that might affect promotions or career progression were also of concern.

Active duty psychologists with more years of active service completed were more likely to express the intention of remaining on active duty until eligible to retire, regardless of the branch of service. In addition, there was a significant relationship between likelihood remain until eligible to retire and years prior active military service completed before becoming a psychologist with those individuals having prior military service being more likely to remain in the service. Career officers with 8 to 10 years of service are more likely to remain until eligible to retire than are officers with less military service completed (Mangelsdorff, 1983, 1984).

In looking at the two major specialty areas (clinician and researcher), the composite profiles of each are revealing in their similarities and differences. The composite clinician rank orders self as a psychologist first and military officer last. The factors likely to affect the clinician remaining until eligible to retire are: promotion eligibility (licensure/certification will be required), the extent the military offers professional growth options, and the availability (or lack of) civilian job openings offering comparable economic and advancement opportunities. The composite researcher is more likely to rate self as a military officer highest. The factors affecting the researcher remaining until eligible to retire include: the extent being a psychologist offers opportunities for advancement and promotion in the military, the years prior active military service completed before becoming a psychologist, the salary paid by the military, and the availability (or lack of) comparable civilian job opportunities. The military offers some distinct advantages.

To retain its military psychologists, the armed services must offer means for continuing professional growth and development. For Army physicians, the availability of residencies and post-residency specialized fellowships has been noted as a significant factor in physician retention (Krause, 1978; Whelan, 1974). The Army offers opportunities for post-doctoral training on active duty (in child, community, or in neuropsychology) for its Army clinicians; a number of Army clinicians are remaining to take advantage of the Army fellowships.

As the services provide the opportunities for professional growth and development, there is also a need for career progressions as psychologists. There has been a limit in the number of senior positions available for the Army clinicians and Army researchers, particularly at the 06 level. The Navy and Air Force appear to have more authorized 06 slots, particularly for the researcher positions. The services need to have distinct career progressions available through the grade of 06, particularly as more psychologists choose to remain to make the military a career.

The armed services can remain competitive in retaining their psychologists by offering: career progressions, opportunities for professional development and personal accomplishment, and options for the development of a professional identity. As more years of active military service are completed (either through training prior service officers as psychologists) or through post-doctoral fellowship training on active duty, the probability increases that the military psychologist will remain in the service until eligible to retire.

REFERENCES

Hedlund, J.L. (1968). Military psychology: A comparative image. American Psychologist, 23, 112-121.

Jacoby, J. (1970). The plight of the uniformed Air Force psychologist. Professional Psychology: Research and Practice, 1 (4), 383-387.

Krause, M.E. (1978). A perspective of motivating factors in the retention of Army orthopedic personnel. Military Medicine, 143, 36-38.

Mangelsdorff, A.D. (1978). Why do some psychologists leave the service? In Proceedings: Psychology in the Department of Defense (sixth symposium). U.S. Air Force Academy, Department of Behavioral Sciences and Leadership.

Mangelsdorff, A.D. (1983). Army physician's assistants: Attitudes, expectations, and satisfaction. Psychology, 20, 1-8.

Mangelsdorff, A.D. (1984). Issues affecting Army psychologists' decision to remain in the service: a follow-up study. Professional Psychology: Research and Practice, 15 (4), 544-552.

Maynard, W.S. (1982). Work activities of uniformed clinical psychologists. An unpublished expansion of a paper presented at the 89th Annual Convention of the American Psychological Association in Los Angeles, California.

Meling, M.A. (1982). Memorandum for Assistant Secretary of Defense (Health Affairs). Survey of the morale of non-physician health care providers (project 2FM-011). Arlington, Virginia: Defense Audit Service report 82-126.

Mitchell, J.L., Keeth, J.B., and Wiekhorst, L.A. (1982). Psychologists in the Air Force: utilization, structure, and job attitudes. In Proceedings: Psychology in the Department of Defense (eighth symposium). U.S. Air Force Academy, Department of Behavioral Sciences and Leadership.

Mitchell, J.L. and Keeth, J.B. (1984). Behavioral Scientists (AFSC 2675), Scientific Managers (AFSC 2616), and Related Specialties. AFPT 90-26X-370. San Antonio, Texas: USAF Occupational Measurement Center, Randolph AFB, Texas.

Murray, J.E. (1978). Air Force clinical psychology survey (USAFSCN 77-151). Washington, D.C.: Personnel and Education Plans Branch, Medical Plans and Health Programs Division, Office of the Surgeon General, United States Air Force.

Rosenbach, W.E. and Gregory, R.A. (1981). Job attitudes of APA military psychology division members. Paper presented at the 89th Annual Convention of the American Psychological Association in Los Angeles, California.

Whelan, R.J. (1974). Why residency programs in the military: The impact of history and influencing forces. Military Medicine, 139, 265-272.

Why Reenlist in the Air National Guard?

Frederick V. Malmstrom
Albert G. Thomas
Jose Gonzales

Washington Air National Guard

ABSTRACT

A locally developed and scored reenlistment survey was administered statewide to 210 randomly selected Washington Air National Guard part-time enlisted personnel. The results suggested that many of the "standard" irritants such as supervisory leadership styles, military uniform standards, and overall policies had no impact on the intent to reenlist. Rather, it was the presence of a meaningful and challenging weekend drill which accounted for most of the intent to reenlist. The results also indicated that, with the aid of microcomputers, it is now quite feasible for local units to develop, administer, and interpret their own professional quality surveys.

BACKGROUND

Employee turnover is an especially difficult variable to predict. Even the best survey instruments seldom have a validity of over r = .5 (R-squared = .25). The intent or the non-intent to remain with a company can be a valid indicator of employee turnover only if the person surveyed feels he or she has complete anonymity; even so, the typical R-squared ranges between .2 to .8. Furthermore, any retention survey has the dual problem of (a) predicting actual turnover, and (b) predicting the intent to remain on the job. It is the second issue this paper addresses.

In mid-1985 the Washington Air National Guard (WA ANG) projected an increase in losses of fully-qualified mid-career (6-10 years of service) part-time personnel. The "normal" annual turnover for the WA ANG was then 14.4%; however, the projected losses in mid-career part timers was estimated to be nearly three times that rate.

In the search for a decent and usable professional survey instrument which could determine why guardsmen and guardswomen would reenlist, it was apparent that none were available for local and state level use. Hence, we were faced with the problems of both developing and interpreting a locally usable survey. It is not surprising that much of the reluctance of local an state units to develop and administer surveys stems from their unfamiliarity with proper survey techniques, including the appropriate statistics. We, therefore, wished to demonstrate these techniques can now be handled quite readily at the local level with the aid of microcomputers and properly instructed laymen.

In conducting such a retention survey, one ought to remember that the Air National Guard has unique features, features not found in the Regular Air Force. First, the ANG part-time force has an unusual mix of civilian occupations and education. For instance, our own enlisted members have a wide spread of civilian skills and ages. Civilian occupations range from attorney and city manager as well to blue collar unskilled and even unemployed workers, all of whom hold hold identical military positions. Furthermore, membership in the ANG is generally more "voluntary" than most laypersons would suspect. Changes in domicile, civilian employment, and "personal circumstances" are frequently sufficient cause to terminate an enlistment. There are wide varieties of reasons for termination of enlistment which the reserve forces accept and the regular forces do not.

METHOD

Subjects. The participants were 210 randomly selected part-time enlisted Washington Air National Guard men and women. The participants were identified in the survey only by sex, age, years of good service, civilian employment, and pay grade. However, for the purposes of analysis, only the categories of years of service (0-3, 3-6, 6-10, and 10-20) were used.

Design. The survey was designed to utilize 20 variables with intent to reenlist as the dependent variable. There were also four optional written response questions, "What I do (or do not) like about the WA ANG," and "Why I intend (or do not intend) to reenlist." The critical group was the mid-career (6-10 year) group; the other year groups served as controls.

Instrument. Because of the unavailability of Air Force surveys for local ANG use, we decided to construct our own. Our survey was tailored from a previously used but unvalidated Puget Sound Power Company "job satisfaction" survey. Basically, our survey sampled nine promising topics as they impacted on the participant's intent to reenlist. They were:

a. Leadership
b. Organizational communication
c. Job recognition
d. Cameraderie
e. Job climate
f. Pay and benefits
g. Promotion opportunities
h. Family support
i. Employer support

A seven-point Likert-type (strongly disagree to strongly agree) scale was employed to rate the 20 question variables which were used for the main body of the survey. The four written responses were also recorded and tallied into a simple like/dislike category.

Procedure. The survey was administered to 210 participants (out of approximately 2,000) part-time personnel over two consecutive weekend drill periods and at six geographically separated locations around Washington state. All participants were guaranteed anonymity.

RESULTS

The results were analyzed by two separate methods. The main analysis on the Likert-type questions was accomplished by multiple stepwise regression on a local microcomputer (Apple IIe). The software program (HSD) had a maximum capacity of 20 variables. Table I shows the cumulative coefficients of determination (R-squared) for the survey questions as they impacted on the intent to reenlist.

Regarding the optional written responses, 95.1% wrote comments on why or why not they liked the WA ANG. However, not more than 10% (n=20) stated why or why not they intended to reenlist; hence, an item analysis of the stated reenlistment reasons was not attempted.

Positive comments regarding the WA ANG outnumbered the negative comments 3 to 1. For the critical 6-10 year group, most of the comments stressed what a positive motivator a dynamic, meaningful training program was, and what a turn-off a poorly organized weekend was. The 0-3 year group tended to stress the importance of pay and benefits, whereas the 10-20 year group tended to stress retirement benefits and personal relationships.

DISCUSSION

Overall, the results of this survey seem both straightforward and unexpected. Unexpected by their absences were variable relating to the perceived leadership and policies of the ANG. Our people did express some strong opinions about these variables, but these variables simply had no impact on their intent to reenlist.

The reasons for the decision to reenlist also seems to vary according to the years of service. That is, there is a distinct break in the reenlistment predictors when one goes "over the hump" at the 10 year service point, a finding also well established in Regular Air Force surveys. Our results suggest that the over-10 year service group would reenlist mainly for desirable job assignments and cameraderie whereas the under-10 year service group would reenlist mainly for a meaningful and challenging weekend. It is noteworthy that the "lure" of pay and benefits tends to become less important to reenlistment with increasing years of service.

Table I

Proportion of Variance Explaining Intent to Reenlist

0-3 YEARS OF SERVICE (n=36)

Question	Cumulative R-squared	p
Pay and benefits are adequate	.372	.001
Family & friends support ANG	.453	.003
Professional military education important	.578	.003
TOTAL POSSIBLE	.617	

3-6 YEARS OF SERVICE (n=46)

Question	Cumulative R-squared	p
I look forward to weekend drills	.566	.001
I would like more paid mandays	.646	.002
TOTAL POSSIBLE	.673	

6-10 YEARS OF SERVICE (n=58)

Question	Cumulative R-squared	p
I look forward to weekend drills	.464	.001
I like my ANG job assignment	.526	.001
I can take my problems to my supervisor	.590	.016
My civilian employer supports the ANG	.623	.014
I understand the ANG promotion requirements	.654	.042
TOTAL POSSIBLE	.664	

10-20 YEARS OF SERIVCE (n=70)

Question	Cumulative R-squared	p
I like my ANG job assignment	.255	.001
My family and friends support the ANG	.371	.001
TOTAL POSSIBLE	.398	

For the 3-6 and the 6-10 year service group, the very high proportion of variance explained by the question, "I look forward to attending weekend drills," is intriguing. The next obvious question is, why do guardsmen and guardswomen look forward to weekend drills? Analysis of the written responses seems to suggest at least two reasons. First, weekend drills, far from being perceived as just another dull and routine military job, represents a discrete break from an otherwise dull and routine civilian job. Second, the majority of negative comments tended to center around a dislike to meaningless and unchallenging training. It was especially irksome to the participants to arrive to a drill weekend (regardless of the pay and benefits) and do nothing. This is equivalent to saying a drill weekend which offers no meaningful training is the best way not to get a guardsman or guardswoman to reenlist.

CONCLUSION

The results of this survey offer several overall suggestions. First, because of the increasing availablity of microcomputers, it is now quite feasible to develop quite respectable retention surveys at the local and unit level. We could, ourselves, write a survey manual, complete with software, so that surveys responsive to the local and unit needs could be administered and interpreted by minimally trained officers and NCO's. Ten or even five years ago this would not have been technologically possible.

Second, most of the retention problems which surfaced in this survey seem to be addressable at the local level. That is, the items local commanders traditionally have no control over, such a National Guard Bureau policy and perceived leadership styles of high ranking officers can be seen to have little or no impact on intent to reenlist. Even the "standard" irritants such a adherence to haircut and military courtesy regulations have little impact on the decision to reenlist in the ANG. This finding suggests that anyone can tolerate these irritants for only two days a month. Rather, we suggest a local unit which offers its people an interesting and challenging weekend will sell itself.

Input Information Requirements for an Adaptive Human-Machine System

P.A. Hancock and Mark H. Chignell
University of Southern California

Abstract

Conventional human-machine task allocation policies are founded upon the assumption that adaptability is intrinsic to the human operator, who may be required to cope with a wide variation in input task loading. Traditionally, the human has been required to compensate for and/or supplement the capabilities of the machine. Developments in artificial intelligence, complex human-machine systems, and methods to assess mental workload imposed by a single task, or group of tasks, allow dynamic task reallocation policies that are responsive to changes in the operatory and task environment. The nature of the information input needed by a Knowledge-Based Adaptive Mechanism (KBAM) for task load leveling is considered. A load leveling KBAM is responsible for implementing task allocation policies which maintain desirable levels of workload on the operator. The problems of how best to assess workload, task-related resource requirements, and resource utilization are considered in the context of the information input required by a KBAM.

Introduction

Background

A key component in the successful operation of contemporary complex systems is the rapid and accurate exchange of information between the machine and its operator. A variety of strategies have been implemented to facilitate such transfer, including improvements in the physical interface and progress in software design relating to concepts such as "user friendliness". Previously, we (Chignell & Hancock, 1985; Hancock, Chignell, & Loewenthal, 1985) have argued that static manipulations of the human-machine interface fail to capture the dynamic essence of this interaction. In response, we have proposed the use of knowledge-based adaptive mechanisms (KBAMs) to improve the level of this interchange. A KBAM is responsive, in real-time, to the dynamic change in overall task demand and reacts by implementing a task allocation policy designed to maintain a satisfactory level of task loading which should in turn prevent unwanted variations in system performance. In the present work, we explore further aspects of this adaptive form of human-machine system and focus specifically upon the derivation of appropriate input information for the KBAM, based upon measures of mental workload (MWL) and attentional capacity.

Task Structuring and Allocation

The type of adaptive interface necessary for task load leveling in human-machine systems will be capable of restructuring the task in accordance with system goals and environmental constraints, and of reallocating task components between human and machine for a given task structure. Figure 1 shows the overall framework for such an adaptive interface. Task structuring would be carried out by a task definition supervisor. As technology advances and human-machine systems develop, the task will no longer be set as a fixed entity. Instead, the definition of the task will change in accordance with the higher level goals set by some external agency. This will also vary with changes in the number and type of environmental constraints acting on the human-machine system. Given a particular definition of the task, allocation of tasks to human and machine would be carried out by a KBAM.

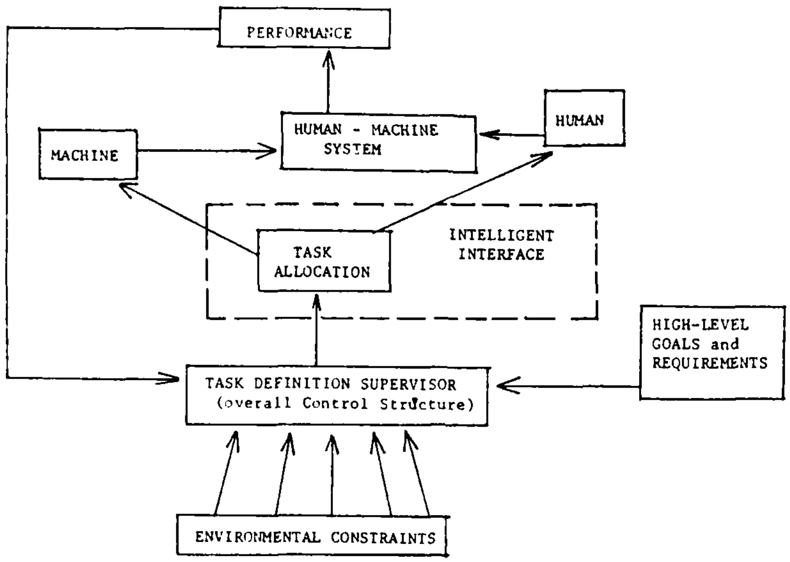


Figure 1. A human-machine system control structure that allows flexible restructuring and reallocation of tasks.

Optimal allocation of task functions between operator and machine components requires a knowledge of human versus machine capabilities. Some of the task functions cannot be involved in the reallocation procedure as they are achievable only by the machine or only by the human component of the system. Allocation is relevant only for tasks which may be switched between human and machine without having a detrimental impact on overall performance. Price (1985) has proposed a method for allocating functions between humans and machines which requires designers to construct a preset and fixed task allocation. However, we suggest that static task allocation policies ignore, in large part, the intrinsic variability in the nature of the task and the human-machine system's response. From the operator's viewpoint, the perceived difficulty of the task is a reflection of the mismatch between task demands and available resources and will vary over time, as illustrated in Figure 2, with the human tolerating the variation in most cases. At times, however, the mismatch may be so great as to produce inadmissible overload or underload which the operator is unable to tolerate, resulting in decrements in performance.

In complex systems, where both task definitions and system capabilities vary over time, task allocation should be viewed as a dynamic rather than static process. In order to achieve a consistently high level of performance, the human-machine system needs to cope with large fluctuations in task demands and operator efficiency. In many cases, the human is able to adapt to mismatches between current task demands and his available capacity. In some situations, however, the mismatch between task demands and available capacity is so great as to preclude sufficient adaptation by the human operator as evidenced in the extremely stressful LANTIRN flight mission in which the human operator appears to be under a state of almost unsupportable overload for large segments of the flight (Roberts & Crites, 1985). At the other extreme, there are a variety of tasks which require the operator to remain relatively passive for long periods, as in the monitoring tasks in oil refineries and nuclear power plants. Practical examples of vigilance present cases of underloading where the KBAM is required to augment task difficulty in order to maintain satisfactory performance (Warm, 1984). In such circumstances, return to the zone of adaptability, i.e., the region where the resources available match the requirements of the task (Figure 2), requires dynamic reallocation of the task components and, possibly, restructuring of the task.

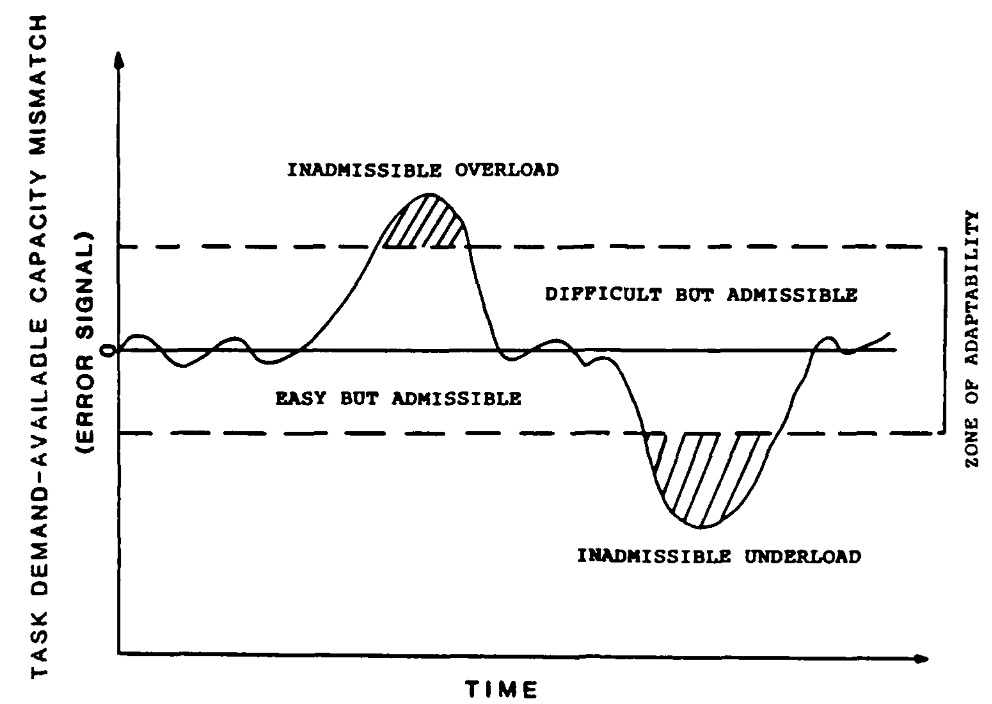


Figure 2. Schematic representation of the time-varying mismatch between task demands and available capacity. Shaded regions indicate inadmissible loading conditions.

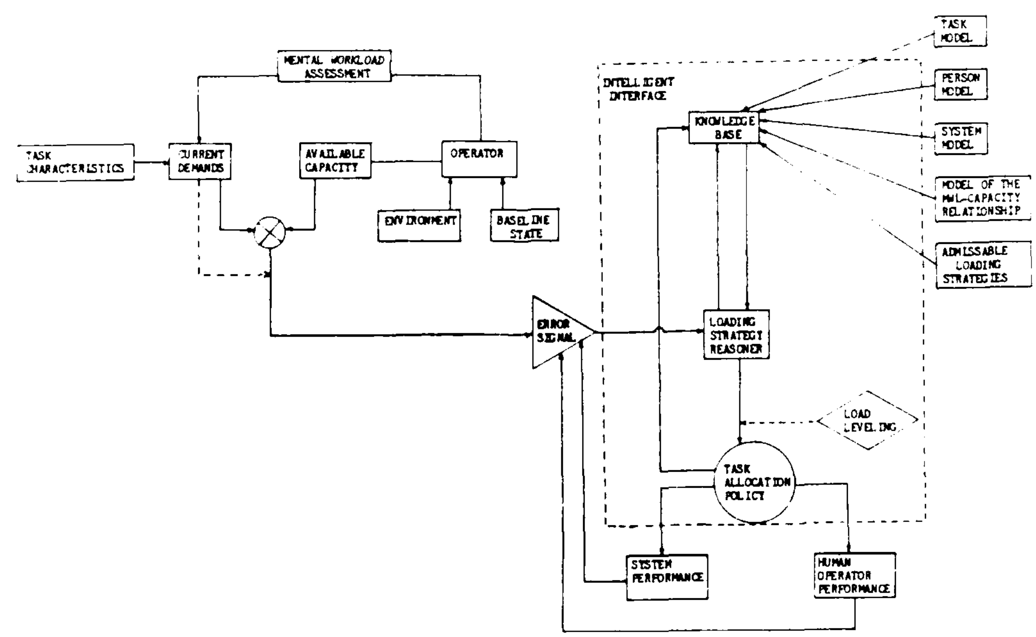


Figure 3. Overall structure of a knowledge-based adaptive mechanism for dynamic task reallocation.

The purpose of a KBAM is to diagnose the state of the machine and operator and reallocate subtasks accordingly so as to optimize performance within the constraints imposed by the properties of the human-machine system and the task environment. There are a number of ways in which a KBAM could be developed. Figure 3 gives an outline of a KBAM design discussed in our earlier work (Chignell & Hancock, 1985).

Information Inputs Required for Taskload Leveling

An adaptive mechanism for task load leveling requires some degree of knowledge-based reasoning in order to deal satisfactorily with the reallocation procedure. The knowledge base required for task load leveling will include details about the complexities of human capability, task structure, physiological variability, and variations in response to imposed load. The first step in developing the KBAM for load-leveling is the identification of appropriate input information to be used by the adaptive mechanism. Ideally, the input to a KBAM will contain information concerning the amount of mental workload (MWL) experienced by the operator, the resources utilized by the human operator, and the resources required by the task, in addition to the usual measures of operator performance. Given adequate information input, the output of the adaptive mechanism will be a definition of the task which alters (where necessary) the loading of task components between the human and machine so as to reduce the mismatch between resources required and resources available. Thus a human-machine interface is developed which acts, initially, as a servomechanism minimizing the difference between current demands and available capacity. The following subsections consider, in brief, how information about MWL, resource utilization, and resource requirements can be acquired. A more detailed review of MWL and resource utilization assessment is given elsewhere (Chignell & Hancock, 1986).

Mental Workload Assessment

Traditional methods of mental workload assessment have employed four major strategies (Moray, 1979). These may be dichotomized initially as subjective reactions to an imposed task load and attentional allocation policies. Under subjective reactions comes the behavioral response of the performer (Vidulich, 1986) and the subjective response as reflected through physiological measures (Hancock, Meshkati, & Robertson, 1985). Each represents the response of the individual and consequently should be calibrated according to each operator. Subjective reactions are vulnerable to the individual's perception and motivation toward the task at hand, but as this is the information often required in a compensatory mechanism, such approaches have practical utility.

The second overall type of strategy depends upon the allocation of attention to the task at hand. The secondary task approach uses the residual capacity which may be apportioned to a task of minor importance as a direct evaluation of attention given to the task of interest. Primary task performance assessment, on the other hand, measures only the actual accomplishment of the operator which, it is proposed, represents a real reflection of attention devoted, except in the case of data-limited tasks (Norman & Bobrow, 1975). Such measures make no inference concerning the quality of attention, merely the amount, so adjustments will have to be made according to the experience and skill of operators and the type of task.

Resource Requirements and Utilization

Recent work in defining and describing tasks has been summarized by Fleishman and Quaintance 1984). The type of description and classification method that is appropriate will depend on the task and the human-machine system. Given a description of the task and an operator with a certain level of skill, however, adequate performance will require a particular level of utilization for each of a number of task-relevant resources. Wickens (1980) has proposed a multiple resource model of attention where different task aspects draw on different pools of attentional resource. This type of model can be used to relate person and task characteristics, allowing a prediction of task performance on the basis of an assessment of the current task and the status of the human operator. Measurement of resources actually utilized by the operator may be possible in future using probe secondary tasks and evoked potential techniques, but at present the assessment of resources being utilized is problematic.

The performance of a task may be thought of as movement through a problem space from an initial state to a goal state. In the means-ends analysis approach to problem solving (Newell & Simon, 1972) procedures are selected according to their ability to reduce the estimated difference between the current state and the goal state. When this approach is applied to human-machine task analysis, each state corresponds to a description of the system at a given point in time. The state space representation forms the basis of the main artificial intelligence techniques for problem solving. The concept of a state space representation of a task can be illustrated using the well known 8-puzzle (Rich, 1984). The 8-puzzle is a square tray in which are placed 8 square tiles. The remaining ninth square is uncovered. Each tile has a number on it. A tile that is adjacent to the blank space can be slid into that space. A game consists of a starting position and a specified goal position. The goal is to transform the starting position into the goal position by sliding the tiles around. From a task-load management perspective, we need a description of the task which includes the notion of amount of mental effort required in each step of task performance. Using the problem space fomulation, it is possible to develop an explicit accounting of task-induced load, at each step of the task. A portion of the state space for a particular 8-puzzle problem is shown in Figure 4. We shall use this figure to motivate our analysis of the resource requirements of tasks in general, which derive, in our view, from four major processes used in carrying out a task.

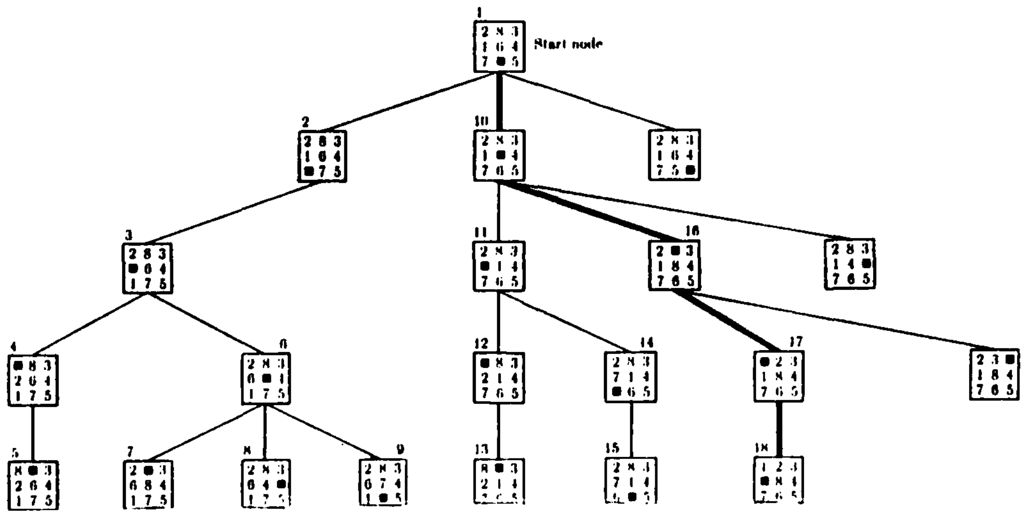


Figure 4. A Portion of the State Space Representation of an 8-Puzzle

Generating moves (node expansion). Solution of the problem requires transition through a number of states which will be referred to as moves in the following analysis. Prior to each move, the person carrying out the task must select a subset of legal moves that can be made. In the case of the 8-puzzle, this selection is easy, since each square adjacent to the empty square represents a potential move. In more realistic tasks, training and experience will determine the ease with which a set of legal (admissible) moves can be generated.

Informational Complexity. The number of moves available, possibly weighted according to the prior probability of selection, determine the informational complexity of the choice (Attneave, 1959). A skilled chess player, for instance, may only consider a small subset of "promising" moves, thereby reducing the informational complexity associated with each move.

Move selection. Once the available moves are known the person then has to choose the best one (in his estimation), i.e. the move which will bring him closest to the goal state. The difficulty of this choice will depend on the discriminability of the best choice from the other choices. If one choice is clearly superior, the

decision is easy. If a number of moves are approximately equivalent in their effect, i.e., they all bring the person equally close to the goal state, the choice becomes more problematic, requiring greater mental effort. The discriminability of a move selection can be quantified using measures related to likelihood, and signal-to-noise ratios.

Action and Subgoal Execution. Once the choice of move is made the preparation o. the action required to implement the choice will require further mental effort. In the case of the 8-puzzle this will include the eye-hand coordination and motor control necessary to move an adjacent tile into the empty square. For high level cognitive tasks, however, this action may in turn be a subtask or subgoal which has its own state space and solution requirements.

Summary

Specification of an appropriate information input is the first step in the development of a working KBAM for task load leveling. Measures of MWL, task-related resource requirements, and resources utilized by the operator appear to be essential parts of this input. This paper considers how these measures can be developed. In our view, study of how resource demands and utilization can be assessed is a prerequisite to adequate understanding of the human component in human-machine systems. The increasing complexity, size, and speed of human-machine systems makes the static view of human-computer interaction inadequate as a framework for new design strategies. The advantages of automatic systems which exclude a human component are offset by the unique knowledge and capabilities brought to bear on the problem by the human, capabilities which can not as yet be emulated in computational processes. Consequently, in order to incorporate the benefits of human capability, without the associated disadvantages, we advocate the use of KBAM methodologies. The potential benefits of the KBAM approach justify a large scale research and development effort. Part of this effort will be an elaboration of the MWL, resource requirement, and resource utilization assessment methods outlined in this paper.

References

Attneave, F. (1959). Applications of Information Theory to Psychology. N.Y.: Holt, Rinehart & Winston.

Chignell, M.H. & Hancock, P.A. (1985). Knowledge-based load leveling and task allocation in human-machine systems. Proceedings of the Twenty-First Conference on Manual Control, Columbus, Ohio, June, 1985

Chignell, M.H. & Hancock, P.A. (1986). A fuzzy comparison of mental workload and available capacity in complex person-machine systems. In W. Karwowski and A. Mital (Eds.), Fuzzy methods and techniques in ergonomics research. N.Y.: Elsevier.

Fleishman, E.A. and Quaintance, M.K. (1984). Taxonomies of human performance: The description of human tasks. N.Y.: Academic Press.

Hancock, P.A., Chignell, M.H., & Loewenthal, A. (1985) KBAM, A Knowledge-Based Adaptive Mechanism, Proceedings of International Ergonomic Conference, Bournemouth, 1985.

Hancock, P.A., Meshkati, N., & Robertson, M.M. (1985) Physiological reflections of mental workload. Aviation and Space Environmental Medicine, in press.

Moray, N. (Ed.) (1979). Mental Workload: Its Theory and Measurement. N.Y.: Plenum.

Newell, A. & Simon, H. (1972). Human Problem Solving. Englewood Cliffs, N.J.: Prentice-Hall.

Norman, D.A. & Bobrow, D.G. (1975). On data limited and resource limited processes. Cognitive Psychology, 7, 44-64.

Price, H.E. (1985) The allocation of functions in systems, Human Factors, 27, 33-45.

Rich, E. (1983). Artificial Intelligence. N.Y.: McGraw-Hill.

Roberts, B.B. and Crites, C.D. (1985). Computer-aided cockpit workload analysis for all weather, multirole tactical aircraft. In Proceedings of the Aerospace Technology Conference and Exposition, Long Beach, CA, October 1985.

Vidulich, M.A. (1986). The cognitive psychology of subjective metnal workload. In P.A. Hancock & N. Meshkati (Eds.), Human Mental Workload. Amsterdam: North-Holland.

Warm, J.S. (1984). Sustained attention in human performance. New York: Wiley.

Wickens, C.D. (1980) The structure of attentional resources. In R.S. Nickerson (ed.), Attention and Performance VIII. Hillsdale, N.J.: Erlbaum.

Cockpit Automation: Passive Versus Active Modes and Their Relationship to State Awareness of the Aircraft

Renate J. Roske-Hofstrand
Aerospace Human Factors Research Division
NASA Ames Research Center

Kenneth R. Paap and Ronald W. Noel
Computing Research Laboratory
New Mexico State University

Abstract

Differences in the type of information stored in memory were investigated as a function of automated versus manual control. Flight parameters displayed on a CRT had to either be monitored and controlled by keyboard entry or simply monitored. On each trial subjects were asked to respond to simulated system failures with a series of keyboard entries signifying recovery procedures. These procedures required the retrieval of information displayed on the CRT before the system failure occurred. Differences between the active and passive mode for the recovery procedures were compared for varying degrees of failure (the amount of system-state information still available) and load (the amount of system-state information required for recovery). The results show that passive monitoring produces greater state awareness when the control task involves low levels of integrality and sequential constraint.

General Project Description

This paper reports the first phase of an extensive investigation that compares an operator's system-state awareness under conditions of active control versus passive monitoring. Many potential benefits can accrue if an operator has constructed an accurate internal representation of the system's current state: decisions concerning what to do next are likely to be more accurate, control actions are likely to be executed more quickly, faults are more likely to be detected, and recovery procedures are likely to be performed with greater speed and accuracy.

The veridicality, completeness, and relevance of the operator's internal representation should be affected by the type of control that the operator is required to exercise. Having to control as well as monitor the environment can be both a blessing and a curse. Active control can contribute to greater state awareness when the motor commands issued during manual control provide an additional memory code that contributes to the construction of the operator's internal representation of the system's current state. Conversely, active control will degrade state awareness in proportion to the processing resources consumed by manual control that are not associated with monitoring the environment, i.e. in proportion to the differences in work load between active control and passive monitoring.

The relative balance between the benefits of the additional motor code and the cost of increased work load will determine the overall difference in state

awareness between the active and passive modes. The goal of this project is to explore two related characteristics of the display, sequential dependencies and integrality, that we feel should enhance the contribution of the motor code to state awareness. If our hypotheses are correct, we would like to develop in a final phase a display that attenuates the loss of the motor code when operator's are engaged in passive monitoring.

Motor codes should be both more relevant and permanent if control actions are governed by a motor schema. Motor schemata are learned sequences of actions that are triggered by specific environmental conditions. The schemata are flexibile enough to permit slots to be filled by values appropriate to the current situation. In contrast to motor schema the term novel action sequence will be used to refer to a sequence of actions that must be generated on-line, one at a time, in response to changes in the environment.

Because motor schema reside in long-term memory they can be activated and remembered as single chunks of information and, accordingly, should be much easier to remember than novel action sequences. Furthermore, since the triggering conditions for a particular motor schema are satisfied by system-state information, these values can be reconstructed or approximated whenever the schema is recalled. Motor programs or schemata are likely to be learned when operators perform control tasks that require stereotypical response sequences. Thus, one of our goals is to systematically manipulate the sequential dependencies in a control task and observe the effect this has on differences in state awareness between active and passive modes.

The second characteristic of control tasks that we will focus on is integrality. In some tasks, like flying, various system parameters can be integrated into a coherent whole. For example, an experienced pilot may be able to encode separate values of airspeed, climb rate, heading, and altimeter and construct an internal representation of the configuration of the aircraft with respect to the ground. However, a pilot manually controlling the plane probably does not continually construct this configuration from the instruments alone. Rather, once an initial configuration is constructed control actions are taken to accomplish successive goals and the control itself serves to time and guide the update and modification of the mental configuration. Accordingly, instruments need only be checked to verify that the expected configuration generated by the last sequence of control actions has been realized. In contrast, pilots engaged in passive monitoring can not benefit from the expectancies generated by control actions.

In a final phase of this project we will "integrate" the individual system parameters into a graphic display. Once subjects have learned how to integrate the system parameters into a coherent configuration, state awareness will be tested both with and without the external, graphic display. Integrality should enhance state awareness under all modes of operation, but when operators must construct the configuration with their mind's eye active control may be essential, particularly under conditions of rapid change.

In summary, our goal is to discover the basic conditions under which the motor code provided by active control is essential to high fidelity state awareness. Achieving this goal will enable one to more accurately predict the desirability of automatic control.

Experiment 1

Method

The goal of the first phase was to develop a basic control task that in its reduced state: (a) contained very little sequential redundancy or integrality, and (b) was very difficult and, therefore, resource demanding. Since these conditions should minimize the effectiveness of the motor code, passive monitoring should generate greater state awareness to the extent that the work load under active control was greater compared to passive monitoring.

Display & Controls. Four instruments were displayed on a color monitor driven by an IBM PC. Each instrument consisted of two side by side rectangles. The slot on the right was labeled "TARGET" and contained a four digit target value that changed from trial to trial, but was constant throughout a given trial. The slot on the left was labeled "ACTUAL" and showed the current value for that instrument. The actual value changed at a rate determined by the forcing function described below. The four displays were labeled AIRSPEED, ALTIMETER, HEADING, and CLIMB RATE, but in this experiment the actual and target values bore no relationship to normal flight parameters.

At the start of each trial the actual value equalled the target value. At various delays the actual values for each instrument would begin to increment or decrement. The rate of change was determined by the sum of three sine functions plus a constant. Intervals between changes were measured in time steps corresponding to 15 video frames of the 60 Hz noninterlaced color display. The intervals ranged from 1 (about .25 seconds) to 18 steps (about 4.6 seconds) with an average of 7.87 steps (about 2.0 seconds).

The actual value would continue to increment (or decrement) until the subject reversed the direction by pressing the opposing button. The F, G, H, and J keys from the second row of a standard keyboard were used to initiate upward shifts in the value of the upper left, upper right, lower left, and lower right instruments, respectively. The keys immediately below (V, B, N, and M) were used to initiate downward shifts. Under optimal control, instruments were incremented or decremented by a value of one. However, each time the subject "over-corrected" by pressing the increment (decrement) key when the values were already going up (down) this would increase the magnitude of each change by one.

Procedure. Each trial consisted of two stages. In the first stage subjects in the active condition performed a tracking task and were instructed to maintain the actual values as close to the target values as possible. An auditory warning signal informed the subject whenever an actual value was more than 10 units from the target. The warning signal persisted until the direction was reversed. The duration of the first stage was randomly drawn from a uniform distribution of time ranging between 64 and 128 seconds.

The second stage of each trial was marked by a half-second blanking of the entire screen that signified a "failure" in the system. When the instruments reappeared one to four of the instruments were marked with a blinking arrow. Subjects were required to use the response keys to indicate if the current value of the designated instruments was above or below the target, i.e. basic system-state information was required to "recover" from the failure. After

the trial was over, feedback was provided by redisplaying all the system-state information and indicating above each of the failed instruments whether the response was correct or incorrect.

There were four types of failures that randomly occurred from trial to trial: (1) both the actual and target values for all designated instruments were missing, (2) the actual value was missing, (3) the target value was missing, and (4) no system state information was missing. The first three failures examine the difficulties associated with retrieving system states while the last indexes the difficulty in recognizing or verifying system states.

When subjects were in the passive mode during the first stage they simply had to monitor the instruments and be prepared to initiate the recovery procedures as soon as the fault occurred.

Design. Four highly trained subjects participated in 10 sessions that lasted about 1.5 hours each. Each subject served as his or her own yoked control: odd-numbered sessions under the active control of the subject were stored in the computer and played back for passive monitoring on the following session.

Each session consisted of 48 trials that contained three replications of each of the 16 conditions formed by the combination of four levels of Load (1-4 responses) and 4 types of Failure. The trials were randomly ordered and, therefore, subjects could not anticipate the amount of system-state information that would still be available during recovery or the number of responses they would have to make. Also, when fewer than four responses were required, the selection of the instruments for testing was equiprobable. One consequence of these randomization procedures was that subjects had to monitor and pay equal attention to all four instruments on every trial.

Results & Discussion

The first two sessions were considered practice and were not analyzed. The remaining eight sessions were partitioned into four blocks with each block consisting of an active session followed by its yoked passive control. Four dependent measures were computed for each trial: (1) proportion of correct responses, e.g. 2 correct out of 3 required responses equals .67, (2) proportion of correct whole reports, e.g. 2 correct out of 3 equals .00, (3) time to the first response, and (4) time to the last response. Each of the dependent measures were submitted to a separate analysis-of-variance that treats Mode (active-passive), Load (1-4 responses), Failure Type (4 levels), and Blocks (1-4) as within-subjects factors.

Proportion Correct Responses. Performance improved with increasing experience ranging from 81% correct in the first block to 92% correct in the final block, $F(3, 9) = 13.84$, $p < .01$. As hypothesized the passive condition produced 7.4% better performance than the active condition, $F(1, 3) = 10.18$, $p < .01$. The active-passive difference did not interact with Blocks or any other combination of factors. The only other significant effect was the main effect of Failure Type, $F(3, 9) = 4.40$, $p < .05$. Performance was about 11% better when both actual and target values were provided compared to the three conditions of missing information which yielded almost identical results.

Proportion of Correct Whole Reports. The critical pattern of results with respect to active control versus passive monitoring were mirrored in the whole report analysis: there was a significant main effect of Mode, $F(1, 3) = 12.45$, $p < .05$ with passive performance (79.7%) superior to active (67.8%). Again, Mode did not interact with any other combination of factors. In contrast, to the previous analysis there was a significant main effect of Load, $F(3, 9) = 16.46$, $p < .001$. Performance declined from 87.2% correct when only one response was required to 63.3% for four responses.

Time to the First Response. Although two of the four subjects were consistently faster in the active condition, Mode did not yield a significant main effect, nor was it involved in any significant interactions. The only significant effect was the main effect of failure type, $F(3, 9) = 8.26$, $p < .01$. When either both values (actual and target) or one value was missing the first response occurred in about three seconds even, but when both values were provided the average time to respond increased by half a second. This is consistent with the view that subjects verified the system state information when they could and that this process takes some additional time.

Completion Times. The pattern of results for time to the last response was nearly identical to that for the first response. The only exception was that completion times also showed a significant main effect of Load, $F(3, 9) =20.58$, $p < .001$. Mean completions times were 3.0, 4.5, 5.2, and 5.0 seconds for one to four responses, respectively. Note that all four responses can be executed just as fast as failures that require only two or three responses. This follows since the situation required subjects to monitor and maintain the state information displayed on all four instruments.

Summary. As hypothesized a control task that is highly demanding and contains very little integrality or sequential constraint generates state awareness inferior to that obtained under passive monitoring. The advantage of passive monitoring over active control is expected to diminish and eventually reverse as integrality and constraint are systemtically added to the basic task.

Acknowledgements

Experiment 1 was supported by cooperative agreement no. NCC 2-206 with NASA-Ames Research Center and the Computing Research Laboratory at New Mexico State University. The authors would like to thank Michael Dotson and Dean Berry for their contributions to the project.

A Comparative Evaluation of Two Subjective Workload Measures: the Subjective Workload Assessment Technique and the Modified Cooper Harper Scale

Dartanian Warr
Human Factors Division, Norton AFB, Calif
And Wright State University, Ohio

Herbert A. Colle
Wright State University, Ohio

and

Gary B. Reid
USAF Aeronautical Medical Research Laboratory
Wright-Patterson AFB, Ohio

Abstract

Twenty-four subjects performed two tasks, a cognitive task and a motor task, both with three levels of task difficulty. Twelve subjects provided workload ratings via the Subjective Workload Assessment Technique (SWAT) and twelve used the modified Cooper-Harper scale (MCH). The objective of this study was to empirically determine if there were differences in the sensitivities of the two subjective workload measures as task difficulty was manipulated. There was no difference between the two techniques' sensitivity. Both rating scales varied significantly as a function of task difficulty manipulations, supporting the sensitivity of both techniques to the workload conditions used.

Introduction

Assessment of mental workload is an important consideration for evaluating alternative system designs. At this time, many system designs are evaluated using performance measures. Performance measures are not always the most sensitive tool, because they sometimes may not give a warning of impending overload problems (Johanssen, Moray, Pew, Rasmussen, Sanders, and Wickens, 1979). Consequently, the use of subjective ratings to estimate mental workload has been proposed as an adjunct to performance based measures (Johanssen, et al, 1979).

To maximize validity, a subjective workload measure needs to be systematically developed. The two subjective measures most often suggested for general mental workload assessment, that meet this criterion, are the Subjective Workload Assessment Technique (SWAT) (Reid, Shingledecker, and Eggemeier, 1981a) and the Modified Cooper-Harper (MCH) scale (Casali and Wierwille, 1982). Both methods have been explored in the literature, which has led to the need to find out if either technique is more sensitive to changes in workload levels. Each technique will be briefly described in the following paragraphs followed by a brief description of the two tasks employed in this evaluation.

The SWAT technique is based on an application of conjoint measurement and scaling procedures (Nygren, 1982) which permit ratings on the

three dimensions (time, mental effort and stress) assumed to be the major contributors to workload (Sheridan and Simpson, 1979; Reid, et al, 1981a). They are then combined into one overall scale of workload which can be demonstrated to have interval properties. In order to identify the rule which is appropriate for combining the three dimensions into an overall interval scale, subjects complete a scale development phase.

During scale development, subjects rank order the 27 possible combinations that result from the three levels of time, mental effort and stress load. This rank ordering information is subjected to a series of axiom tests to identify the rule for combining the three dimensions. When the rule has been established, conjoint scaling is applied and the appropriate scale for workload is derived. A Kendall's coefficient of concordance is calculated for the subjects' rank orderings. If the result is .79 or higher then the scale which is developed will be used.

The MCH scale was developed based on the Cooper-Harper (CH) (1969) aircraft handling qualities rating scale. The CH was developed to help evaluate aircraft flying qualities during aircraft flight testing and development. The CH was found to both valid and reliable in its specified task. Further evaluation, also found that the scale used many words as anchors which are used to describe operator workload (Moray, 1982; Wierwille and Williges, 1979). Wierwille and Williges (1979) suggested that if the CH was modified to further describe operator workload then this new scale could be used as a subjective workload measure. In 1982, Wierwille and Casali modified the CH, producing the Modified Cooper-Harper (MCH) scale. It is a ten-point scale which uses a logic tree to help the users rate workload. The ten points are anchored with very specific descriptors which help to provide consistent ratings.

These techniques were compared using two tasks, a motor task and a cognitive task, each with three difficulty levels. The motor task was an unstable tracking task with its difficulty levels provided by changing the stability in the tracking system. The cognitive task was the continuous recall. In this task, pairs of numbers are presented vertically on a monitor. The subject memorizes the bottom number and decides if the top number is the same as the bottom number presented 1, 2 or 3 screens earlier. The difficulty of this task is manipulated by the memorization task and the number of digits in a number, either 1, 2 or 4 digits. Both tasks were taken from the Criterion Task Set (CTS) developed by Shingledecker (1984) for use in the development of workload measures.

Method

Twenty-four subjects, (12 men and 12 women) enrolled in introductory psychology courses at Wright State University, received extra credit for their participation in this study.

Apparatus

The two tasks were presented on a 12 inch black and white monitor which was controlled by a Commodore 64 computer. Subjects sat approximately 2 feet from the monitor and used either a control knob for the tracking task or a pushbutton pad for the recall task to respond to the system.

Procedure

Subjects were randomly assigned to the scale groups. 12 Subjects were trained in MCH use and 12 were trained in SWAT use, with men and women equally distributed throughout the groups. Once the subjects were trained on their scales, they were then trained on the two tasks. Each subject received 12 practice trials, 6 on the recall task and 6 on the tracking task. For each task, 3 trials were at the low difficulty level and 3 were at the high difficulty level. The first two practice trials were at the CTS practice level while the last practice trial was at the CTS test level (Shingledecker, 1984). After training was completed the subjects began testing.

Each subject received 12 test trials. After each trial, the subjects rated the workload. These workload scores were recorded by the researcher. Also, four performance measures were taken. For the tracking task, the root mean squared error (RMS) and the number of control losses were recorded. For the recall task, the mean reaction times and the percent of incorrect responses were recorded.

The ratings generated by these trials were analyzed using a 2 (scales) by 2 (tasks) by 3 (difficulty) mixed factorial design. The first factor was scales (SWAT vs MCH). It was a between-subjects factor. The other two factors were task type (tracking vs recall) and task difficulty (low, medium and high). These were repeated-measures factors. All subjects performed both tasks and received all levels of difficulty for each tasks.

The task difficulty combinations and order of task presentation were balanced to control for the effects of practice and fatigue. The experimental order provided each subject with 6 recall trials and 6 tracking trials.

A linear transformation (TSWAT= .09SWAT + 1) was performed prior to the analysis of variance (ANOVA), to make the TSWAT scores equivalent to the MCH scores. This transformation of SWAT scores is permissable, if the SWAT scale developed has met the axioms which validate the SWAT as an interval level scale. A Kendall's coefficient of concordance on the SWAT scale development was .85 which allowed the use of the SWAT scale developed and the SWAT scores.

Results

The three way ANOVA found no scale interactions statistically significant; scales by tasks, $F(1,22) < 1.0$, scales by difficulty, $F(2,44) < 1.0$, and scales by tasks by difficulty, $F(2,44) < 1.0$. Thus, there is no evidence of the scales differing in sensitivity at the .05 level.

The main effect of task difficulty was statistically significant, $F(2,44)= 44.5$, $p < .01$, which indicates that both scales where sensitive to the task difficulty manipulations. However, the interactions describe above indicates that the two scales did not differ in sensitivity. The mean ratings from both scales are presented in Figure 1. The left panel presents the ratings from the recall task and the right panel presents ratings from the tracking task. As Figure 1 shows, the ratings from the scales increased substantially as task difficulty increased. For the recall task, the mean ratings were 5.6, 6.5, and 8.3, for the low, medium and high difficulty levels respectively. For the tracking task, the mean

ratings were 2.9, 6.9, and 8.8, for the low, medium and high difficulty levels respectively. The other main effects of scales, $F(1,22) < 1$ and tasks, $F(1,22) < 1$ and the interaction of tasks by difficulty, $F(2,44) < 1$, were also found to be statistically insignificant.

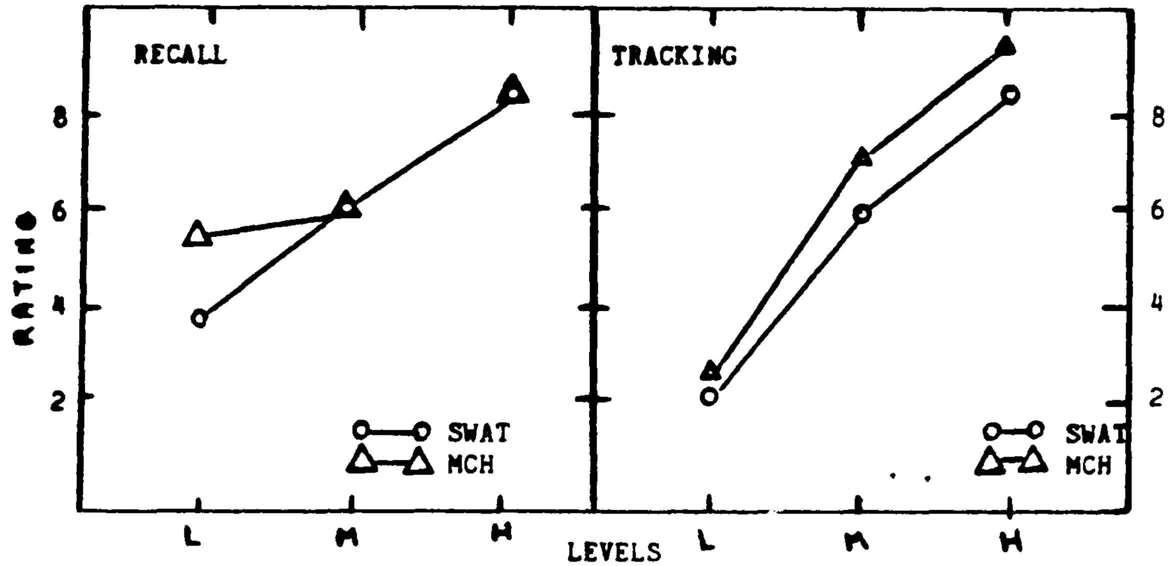


Figure 1. Subjective ratings as a function of task difficulty manipulations on two types of tasks.

During this study, four performance measures were recorded, two for the recall task (mean reaction time and percent of incorrect responses) and two for the tracking task (RMS and total control losses). Each of the four measures was analyzed separately using a two way ANOVA. The scales (MCH and SWAT) were the between-subjects factor, while the task difficulty levels (low, medium, and high) were the repeated-measures factor.

Each performance measure had a main effect for the task difficulty levels. The task difficulty main effect for the recall performance measures, mean reaction time [$F(2,44)= 140.5$, $p < .01$] and percent of incorrect responses [$F(2,44)= 11.9$, $p < .01$], both were statistically significant. The main effect of task difficulty was also statistically significant for the tracking task performance measures, RMS [$F(2,44)= 163.7$, $p < .01$] and total number of control losses [$F(2,44)= 214.9$, $p < .01$]. These results indicate that each performance measure was sensitive to the task difficulty manipulations. There was no interaction with scales found for any performance measure, which indicates that the tasks were equivalent for both scale groups. Nor was there a main effect for scales found for any performance measure.

The performance measures' mean scores, for each task difficulty level (low, medium and high), are as follows: reaction time 678.6, 978.6 and 1276.7, respectively; percent of incorrect responses, 31.8, 39.0 and 48.0; RMS, 13.8, 36.5 and 36.7; and total number of control losses, 8.0, 213.9 and 410.5. These scores are comparable to those found by Shingledecker (1984).

Discussion/Conclusion

These results indicate that SWAT and MCH ratings are comparable and are equally sensitive to variations in task difficulty. Also, these results indicate that both subjective workload measures are sensitive to difficulty manipulations for motor and cognitive tasks, which suggest that these measures maybe equally sensitive to a wide range of tasks. Although the scales were found to be comparable in the present setting, it is not clear if these results would be repeated in an applied setting, where the operators would be more familiar with the tasks and the expected difficulty levels. If the MCH and SWAT techniques are shown to be comparable in a number of settings, then such factors as ease of use, intrusiveness and operator acceptance might be the next areas in which to explore the advantages of these two techniques. Also, in some situations, the diagnostic information obtain from the individual SWAT scales might be of interest. In the current study, only the overall SWAT scores were compared with the MCH scores.

References

Cooper, G.E. and Harper, R. P. (1969). The use of pilot rating in the evaluation of aircraft handling qualities. Moffett Field, California: National Aeronautics and Space Administration AMES Research Center, Report, TN-D-5153.

Johannsen, G., Moray, N., Pew, R., Rasmussen, J., Sanders, A. and Wickens, C. (1979). Final report of Experimental Psychology Group. In N. Moray (Ed.) Mental Workload: Its Theory and Measurement. New York, New York: Plenum Press.

Moray, N. (1982). Subjective Mental Workload. Human Factors. 24,25-40.

Nygren, T. E. (1982). Conjoint measurement and conjoint scaling: User's guide, Wright Patterson AFB, Ohio: Air Force Aerospace Medical Research Laboratory, Technical Report AFAMRL-TR-82-22.

Reid, G. Shingledecker, C. A., and Eggemeier, F. T. (1981a). Application of conjoint measurement of workload scale development Proceedings of the 1981 Human Factors Society Annual Meeting. 522-526.

Sheridan, T. B. and Simpson, R. W. (1979). Toward the definition and measurement of mental workload of transport pilots. Cambridge Mass: Massachusetts Institute of Technology Flight Transportation Laboratory Report, FTL Report R 79-4.

Shingledecker, C. A. (1984). A Task Battery for Applied Human Performance Assessment Research. Wright-Patterson AFB, Ohio:

Wierwille, W. W. and Casali, J. G. (1983). A validated rating scale for global mental workload measurement application. Proceedings of the 27th Annual Human Factor Society Meeting, 129-133.

Williges, R. C. and Wierwille, W. W. (1979). Behavioral measures of Aircrew Mental Workload. Human Factors, 21, 549-574.

An Assessment of the Tactical Benefits Afforded by Voice Actuated Controls in a Single Seat Fighter: Methodology and Some Preliminary Results

Mr. Brian Melville
Essex Corp., Fort Rucker, Alabama

2Lt Jay Horn
Crew Station Design Facility, Wright Patterson AFB, Dayton Ohio

Mr. Christopher Goulet
Human Factors Branch, Wright Patterson AFB, Dayton Ohio

Abstract

One solution for managing the high level of information transfer between the pilot and cockpit of a future single seat fighter may be through voice actuated controls. Speech recognition technology needs to mature before it can be effectively placed in the cockpit; however, a need exists now to ascertain the benefits voice control would offer in the future tactical environment in order to help determine the level of funding of speech recognition research. This study assessed whether voice actuated controls improved aircraft survivability to kill ratio and reduced pilot workload over current manually actuated controls. It utilized a high fidelity simulator representing the advanced fighter cockpit of the near future and full mission scenarios representative of high and low workload tactical situations. Preliminary results of the 8 participating pilots' subjective opinions and of the amount of time the pilots spent looking out of the cockpit at the surrounding tactical environment while reconfiguring weapons, indicated that segment of a mission where critical tasks are to be performed in a limited and compresssed amount of time may be accomplished more easily, safely, and effectively using voice actuated controls. The results also indicate that voice actuated controls may be easier and safer for a wingman to use while following a lead aircraft in adverse weather, at night, or in terrain following flight. A more in-depth analysis of the data is needed before definitive conclusions can be reached.

Introduction

Background

With the advent of computerized speech recognition, vocal commands can now direct computers to perform controlling functions. This has enabled operators to communicate with computer systems in a manner most natural to them, talking. Diverse applications of voice control, ranging from personal computers to large material handling processes, have achieved success. These successes have come largely from the operator's capability to use the eyes and hands for other tasks. The complex cockpits of future tactical fighters may be another area where computerized speech recognition and voice control may save operator workload and enhance system performance.

Recent trends in cockpit design, as described by Warner and Harris (1984), are toward decreasing the physical size of the crew station while incorporating increasingly sophisticated weapons and sensor systems. The result is less panel space for controls and displays and an increasing amount of information that must pass between the pilot and his aircraft. In addition, the majority of fighter aircraft are not single-seat, thus assistance from a second crewmember, who in the past served as a weapons officer, is no longer available. As a solution, integrated controls and displays and multi-function control and display systems (MFC/D) have been developed allowing the same cockpit space to control several different subsystems.

Warner and Harris (1984) point out that MFC/D systems have met the requirement of increased information transmission but not without creating new problems, specifically; 1) control actions that are simple and direct in a conventional cockpit are made considerably more complex when the pilot must first access the desired subsystem through menues on his control display set, and 2) spatial and tactile cues previously available to aid the pilot in locating and operating subsystems are absent in MFD/D operations placing an additional burden on the visual and manual information channels. Saturation of the visual and manual channels is compounded as modernized enemy threats require the pilot to fly increasingly lower and faster. It is not hard to see that the visual and manual channels of today's fighter pilot are overloaded.

Voice control may be one solution for managing the complex interaction of information between the pilot and aircraft. Research indicates an advantage of voice over manual control for certain cockpit functions (Aretz, 1983, Williamson and Curry, 1984, North, Mountford, Edman, Guenther, 1980) and recognizers capable of limited operations in the fighter environment have been demonstrated. However, Flanagan, Dixon, Doddington, Makoul, McCauley, Roland, Ruth, Simpsons, Williges, Woods, and Zue (1984) report the current level of speech recognition technology unable to reliably operate in such high stress environments and that a great deal of research needs to be accomplished before this capability can be attained.

Purpose

The purpose of this study was to assess the tactical utility of voice control in the context of a full mission scenario by comparing current manual push-button control with voice control on the basis of aircraft survivability to kill ratio and pilot resource loading and workload. The study was conducted to provide data which may influence the funding of speech recognition research by demonstrating whether or not voice actuated controls can provide tactical benefits.

Approach

The weapons set up and delivery phases of an air-to-ground mission were selected as representing; 1) general piloting tasks that are normally accomplished in low to moderate workload situations under little time criticallity (such as radio frequency changes, IFF code changes), and 2) high workload time critical mission segments involving a variety of complex tasks demanding simultaneous attention, compressed time and high stress. To assess the future utility of voice control it was necessary to simulate a future level of speech recognition accuracy. To accomplish this the performance of the speech recongizer was artifically enhanced by restricting the number of speech commands to discriminate amongst and by selecting commmands which offered the greatest degree of phonetic dissimilarity. The speech commands and weapons systems controlled were as follows:

NAV MODE	AIR TO GROUND		(MASTER MODE)
MARK 82	MAVERICK	GUNS	(WEAPONS)
POWER ON	BORE	SLAVE	(MAVERICK SET UP)
MANUAL	CCIP	DIVE TOSS	(DELIVERY MODE)
PROFILE 1	PROFILE 2		(DELIVERY PROFILE)
SINGLE	PAIRED		(RELEASE)
50 FT	80 FT	100 FT	(INTERVAL)
3	5	7	(PULSES)
NOSE	TAIL	FUSE BOTH	(FUSING)

On the basis of a survey of pilot opinions regarding what cockpit systems should and should not be controlled by voice (Canyon study), the actual "arming" and "release" of weapons was always under manual control.

Two experiments were conducted in this study. The first utilized missions that were as representative of the real world as possible and thus directly addressed the issue of ascertaining tactical benefits afforded by voice control. The second experiment used missions which unrealistically drove the pilot to task saturation and comparatively explored the limits of voice and manual control. Time and space constraints dictate that only the methodology and some preliminary results of experiment one be presented here.

Methods

Subjects

Eight current F-16 pilots, averaging 29 years of age and six years military experience, served as test participants. Flight hours ranged from 300 hours to 2400 hours with a mean flight experience of 950 hours. All pilots listed air-to-ground as their primary background and none of the pilots reported combat experience.

Apparatus

The study utilized the F-16 full mission simulator residing at the Crew Station Design Facility, Wright Patterson AFB. The F-16C was a fully instrumented fixed-base simulator modified for speech recognition research. These modifications included: 1) incorporation of a fourth position on the communication transmit button located on the throttle, for the pilot to toggle the voice control system on and off, 2) the addition of a partitioned area in the lower left portion of the heads up display (HUD), to provide visual feedback on the status of the voice control system (this area was outlined with a small rectangular box anytime the voice control system was activated by the pilot and "listening" for commands), and 3) incorporation of a Shure SM10 microphone, used for communicating with the speech recognizer, to the standard H-157A/A1C boom mic and headset. HUD symbology, which the pilot used for aircraft orientation, navigation, and weapons delivery, overlaid a visual scene of moving terrain. The visual scene represented 450 square miles of eastern Pennsylvania and was projected from a moving terrain belt, as a function of aircraft altitude, heading, and ground speed. The simulator was driven by six Systems Engineering Laboratory computers which housed the basic aero, weapons control, fire control computer, and navigation models was used for data collection. The experimenters monitored the simulator from a console and directed the test participant's flight via headset by playing the role of Forward Air Controller (FAC). Manual push-button control of the weapon systems was accomplished using the buttons surrounding the F-16's Multi-Function Displays (MFD). A Texas Instruments Portable Professional Computer containing a discrete speech recognition board was used

to control the weapons multi-function displays by voice. Video and audio recordings of the pilots were obtained with a RCA model TC1020 camera and recorder, a TEAC video recorder, and reviewed on a Sony model BVE-500A editing machine.

Missions

Four missions were developed with input from the operational community to each include one weapons setup segment and one weapons delivery segment of I2R Maverick missiles or MK 82 bombs. In the setup segment the maverick missile's seeker heads were cooled (readied) by selecting "Maverick" and "power-on" through menus on an MFD or by voice command. The setups were accomplished either at high altitude while orbiting or while terrain following at 400 knots and 500 feet above the ground.

Weapons delivery segments were conducted from a low level (500 ft) terrain following altitude and always included a last minute change in either target type of target orientation which forced the pilot to change the type of weapons he was delivering or his delivery profile. In one mission for example, 45 seconds before the target, the pilot was informed by the FAC that his primary target of tanks were destroyed and to hit his secondary target of troop trucks, thus the pilot had to switch from maverick anti-tank missiles to general purpose MK 82 bombs. This involved one MFD button press or one vocal command. In another mission, the pilot was informed 45 seconds out that the column of trucks that were his primary target had dispersed with the FAC recommending the following delivery profile; "CCIP", "Paired" release, "50 ft" spacing, "5" pulses and "Tail" fusing. This involved ten MFD button presses or five vocal commands while inbound to the target.

During the weapons delivery segment of the four missions, the pilot had to: 1) maintain separation from the ground, 2) precisely fly the briefed route of flight to avoid being shotdown by enemy threats, 3) respond to enemy threats, which automatically appeared regardless of how accurate he maintained his route of flight, on his threat warning display by dispensing chaff and flares 4) change weapons or delivery profile, and 5) visually acquire the target and position the aircraft for accurate delivery (MK 82s) or slew and designate the maverick missiles. Radio squeals, simulating extraneous communication,were also piped into the pilot's headset during the segment to add to the confusion.

The weapons load for the four missions were 8 MK 82s and 2 maverick missiles. It was realized that in "real world" air-to-ground missions, a pilot would not change weapons or delivery profiles in the target area, or carry a mixture of air-to-ground missiles and bombs. Obviously, something had to be changed in order to compare voice and manual control, and reconfiguring weapons while in the target area and carrying a mixture of weapons, may not be entirely unfeasible in the future.

Measures

The measures were broken into areas of survivability, weapons delivery, pilot resource loading and workload. The survivability measures included the number and total duration of exposures to enemy threats, flying into the bomb fragmentation pattern as a result of failing to change to tail fusing, the adequacy of the chaff/flare releases, the average speed through the target area (faster is better), and the average above-ground-altitude through the target area (lower is better). Weapons delivery included the number of weapons released and the number of kills (accuracy). Pilot resource loading ard workload were measured by the amount of time the pilot spent looking through the HUD, the time his hands remained on the stick and throttle, the number of chaff/flare releases, the amount of time and the accuracy of reconfiguring his weapons, and by Subjective Workload Assessment Technique (SWAT) ratings (Reid, Shingledecker, and Eggemeier, 1981). Subjective comparisons of voice and manual control were made using a 5 point scale on; 1) which was more useful in controlling the MFDs, 2) which offered higher operational utility (i.e., higher kill to survivability ratio or ability to stay with lead aircraft if flying wing, and 3) which was best in reducing pilot workload.

Procedures

The experiment encompassed two two-factor factorial designs each comparing voice and manual controls on different types of piloting tasks, weapons setup (representing general piloting tasks), and weapons delivery (representing tasks conducted in a high workload-time critical situation). The first was a repeated measures design exploring voice vs manual control at low vs high degree of system interation. The second was a between subjects design and assessed voice vs manual control on general piloting tasks at high vs low altitude.

Pilots arrived in pairs for three days of participation. The inhouse training and familarization was expedited by sending each pilot a week in advance, a phamphlet explaining the background and purpose of the study, schedule, basic cockpit layout and operation of the F-16C including preflight checklists, terrain following flight, and weapons delivery, a description of SWAT, and a description of the theory behind speech recognition and the procedures for enrolling their individual speech characteristics into the voice recognition system. The first day was devoted to developing the pilots' individual SWAT workload scales, as described in Reid et al. (1983), having the pilots enroll the 24 word vocabulary into the speech recognition system, and to training and attaining a comfortable level of flying low level terrain following missions, delivering maverick missiles and MK 82 bombs, and changing weapons and profiles prior to delivery

using both voice and manual methods. The training utilized three missions developed for this purpose.

The four data collection missions were flown on the 2nd and 3rd day in an order such that manual versus voice control was balanced within subjects, the degree of system interaction of the weapons delivery segment (weapon change) vs profile (change), was balanced between subjects, and maverick setup segments (high altitude, orbiting vs low altitude, terrain following) balanced across subjects.

For each mission, the pilot was briefed on a primary and secondary target, the route of flight, and the location and nature of enemy threats along his route of flight. The pilot was then directed to configure his weapons in a format specified by the experimenter to ensure that each participant had the same delivery profiles at the start of a mission and hence the same number of changes to make in the target area. The initial configuration of weapons was accomplished by the pilot using the type of control he would use during the mission, manual or voice, in order to familiarize him with how his weapons were configured and as a further reminder of what type of control he was to use during the mission.

In flying the mission, the pilot followed the briefed flight plan, setup his maverick missiles upon crosssing a specific checkpoint, changed his weapons or delivery profile at the direction of the FAC, and delivering his weapons to the assigned target. Each mission was scripted with communications threat warnings (other than those caused by straying off the flight path), and radio squeals, cued by the computer to ensure that all radio communications and distractions were uniform in timing and content. The computer also activated data collection during the weapon setup and delivery segments automatically a a function of aircraft position so that aspects of the mission extraneous to the segments where voice or manual control was being exercised would be excluded from the data. After weapons delivery was completed, the mission was over, and the pilot completed a SWAT evaluation sheet, recorded his perceived level of workload during: 1) the maverick setup and, 2) the weapon delivery phase. Upon completing all missions, each pilot completed a questionnaire..

Results

On the basis of time constraints, only the amount of time the pilots spent looking through the heads up display, and their subjective opinions as registered on the questionnaire, are presented. A repeated measures two factor Analysis of Variance (ANOVA) was conducted on the "head up" measures for the weapons delivery segments and disclosed a significant difference between voice and manual control $F(1,28) = 16.28$, with voice offering almost 10% more time heads up (80% manual vs 89.4% voice). There was no significant main effect of weapon vs profile change, F 1, or interaction, F 1. A mixed two factor ANOVA was used for the "head up" measures on the maverick setup segments and found no significant main effect of voice versus manual control. $F(1,12) = 2.06$, high vs low altitude, $F<1$, and no significant interaction, $F<1$.

Subjective comparisons of voice and manual control on: 1) how useful they would be to control multi-function displays, 2) how they would enhance operational utility (i.e., improve survivability to kill ratio, or make it easier to fly as a wingman), and 3) how effectively each would reduce pilot workload, were made by the pilots with regards to the generic situation that the mission segments representd. Specifically the weapons delivery segment represented high workload time critical situations with high and low degrees of aircraft subsystem interaction and the maverick setup segment represented the execution of general piloting tasks while flying as a wingman at high altitude and at low level while terrain following. The pilots made projections to the "real world" on the basis of their experiences in the study. The responses to these questions are presented in figures 1 and 2.

Discussion

The fact that there was a 9.4% increase in the amount of time the pilot looked outside his aircraft confirms the postulated benefit of voice control of freeing the eyes for more important tasks. This additional 5.6 seconds per minute of heads up time could mean a lot to a pilot having a critical need to keep abreast of the surrounding tactical environment. This is evidenced by the pilots' opinions of the superiority of voice over manual control on usefulness, operational utility, and reduction in pilot workload during high workload, time critical situations involving a high degree of subsystem interaction and terrain following flight that requires following a lead aircraft. On the basis of the responses to flying as wingman at high altitude or in situations involving a low degree of system interaction, it appears that benefits afforded by voice control decrease as the amount and criticality of tasks decrease. However, a substantial amount of data in this study has yet to be examined before definitive conclusions can be drawn on the benefits of voice actuated control to a tactical scenario.

References

Aretz, A.J. (1983) A Comparison of Manual and Vocal Response Modes for the Control of Aircraft Systems. Proceedings of the Human Factors Society 27th Annual Meeting

Cotton, John C., McCauley, Michael E., North, Robert A., and Strieb, Melvin I.

(1983), Development of Speech Input/Output Interfaces for Tactical Aircraft AD-A136485 (Canyon Study).

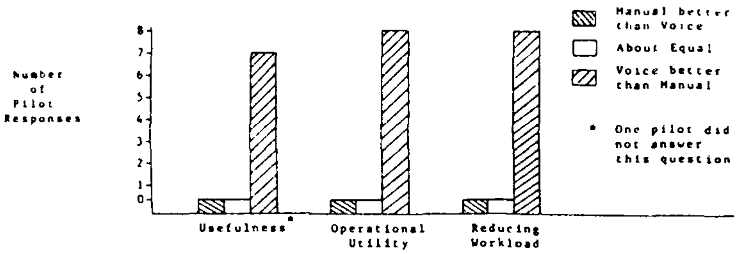
Flanagan, J.L., Dixon, N.R., Doddington, G.R., Makoul, J.I., McCauley, M.E, Foland, E.F., Ruth, J.C., Simpson, C.A., Williges, B.H., Woods, W.A., and Zue, V.W. (1984) Automatic Speech Recognition in Severe Environments. Washington, D.C.; National Research Council

North, R.A., Mountford, S.J., Edman, T, and Guenther, K (1980) The Use of Voice Entry as a Possible Means to Reduce Fighter Pilot Workload: A Practical Application of Voice Interactive Systems Voice-Interactive Systems: Applications and Payoffs 253-263

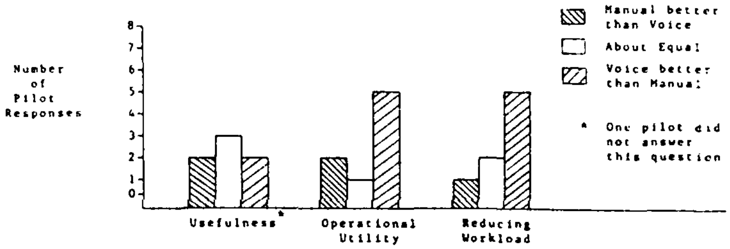
Reid, G.B, Shingledecker, C.A., and Eggemeier, F.T. (1981), Application of Conjoint Measurement to Workload Scale Development. Proceedings of the 1981 Human Factors Society Annual Meeting, 522-526.

Warner, N., Harris, S. (1984), Voice-Controlled Avionics: Programs, Progress and Prognosis. Speech Tech '84 110-123

Williamson, D.T., Curry, D.G. (1984), Voice I/O Effectiveness Under Heavy Task Loading. Journal of American Voice I/O Society

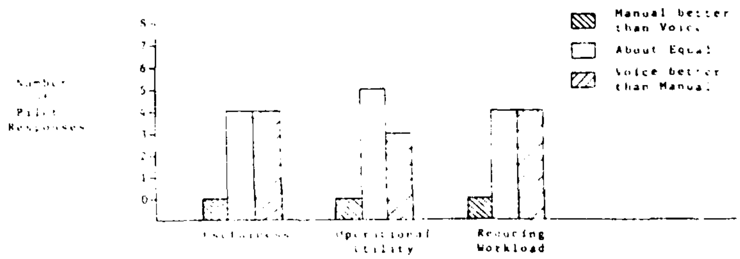


(a) High Degree of System Interaction

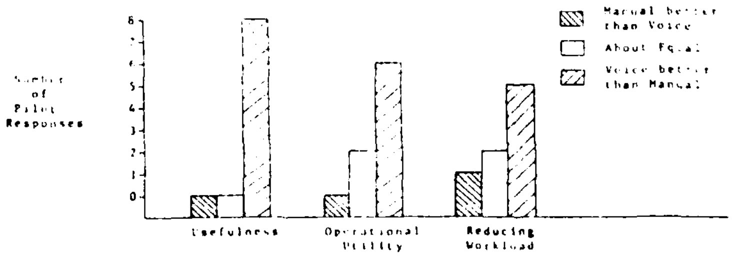


(b) Low Degree of System Interaction

Figure 1 High Workload Time Critical Situations



(a) Flying as Wingman at High Altitude



(b) As Wingman during Low-Level Terrain-Following Flight

Figure 2 Executing General Piloting Tasks

A COMPARISON OF TWO SUBJECTIVE FATIGUE CHECKLISTS

J.C. Miller and A. Narvaez
Crew Technology Division
USAF School of Aerospace Medicine
Brooks AFB, Texas

INTRODUCTION

The Crew Performance Laboratory of the USAF School of Aerospace Medicine's (USAFSAM) Crew Technology Division has two subjective fatigue checklists available for use in operational environments. We undertook the investigation described here to provide comparative information about the relative sensitivities and reliabilities of the two checklists. We studied two groups of subjects: office workers and USAF transport aircrew members.

METHODS

The subjects in this investigation completed two different rating checklists designed to estimate subjective fatigue. The Subjective Fatigue Checklist (20PT), also known as the "Feeling Tone Checklist," was developed at the USAFSAM in the 1950's (Pearson and Byars, 1956; Martin and Venables, 1980). It is a unidimensional scale of equal-appearing intervals (based on expert judgment) ranging from "Very Refreshed" to "Extremely Tired" (Storm, 1983). For each of ten reliable, sensitive statements, the subject determines whether he or she feels "2 - Better than," "1 - Same as," or "0 - Worse than," the feeling described by the statement. This results in a total integer score ranging from 0 to 20 with lower scores indicating greater fatigue. We interpret the subjective fatigue scores in terms of both relative values and absolute scores. In general, average subjective fatigue scores of 12 and above suggest feelings of alertness; 11 down to 8, moderate fatigue; and 7 and lower, severe fatigue (Storm and Merrifield, 1980). The checklist is sensitive to 24-hour fluctuations in fatigue (Friedmann et al., 1977).

The second form, the 7-point checklist (7PT) was developed during the period, 1978-80, by the staff of the USAFSAM Crew Performance Branch, principally L.P. Perelli and W.F. Storm. The newer checklist was developed in response to aircrew members' difficulties understanding some 1950s colloquialisms in the 20PT and aircrew needs for a checklist that could be completed more quickly than the 20PT. Perelli and Storm selected the statements anchoring the checklist points through iterative presentations of draft versions of the checklist to aircrew members. The subject circles the number of one of the seven statements that best describes how he or she feels at the moment.

<u>Office Workers</u>. A repeated measures design was used to collect fatigue ratings from office workers. They were employed at a local university or at a local Air Force base. Occupations

included clerk typists, secretaries, data entry clerks, and administrative assistants. Twenty-nine female workers and one male worker participated. They were told that we would compare two forms which had been developed to measure subjective fatigue; that the information gathered would be used to calibrate data already collected from aircrews; and that the relationships among the forms would be examined.

The order of presentation of subjective fatigue checklists was counter balanced. Each worker always completed the two forms in the assigned order. The worker received a packet containing a summary of the nature and purpose of the study, instructions on how to complete each checklist, and four stapled sets of the two forms. She or he completed the two forms four times throughout the course of one day: upon arrival at work in the morning, before leaving for lunch, upon returning from lunch, and before leaving work at the end of the day. One week after the date of this first participation, the workers repeated the same procedure (on the same day of the week).

Nonparametric statistical analyses were performed with the Biomedical Statistical Package program, BMDP3S (Dixon, 1981), with reference to Siegel (1956). Sample median ratings ($n = 30$) for the 7PT and 20PT were calculated for each of the four periods on each of the two days. The Friedman two-way analysis of variance by ranks was used to investigate median differences within days. Medians for day 1 and for day 2 were calculated from the four observations per day per worker. The differences between day 1 and day 2 sample medians ($n = 30$) were examined with the Wilcoxon matched-pairs, signed-ranks test (2-tail). Statistical significance was accepted beyond the 95% level of confidence ($p < .05$).

The correlation ($n = 30$ workers) between median day 1 and median day 2 subjective fatigue ratings was estimated as the Spearman rank correlation coefficient. The intercorrelation ($n = 240$ observations) between the 7PT and 20PT was assessed using the same coefficient. We calculated inter-rater agreement ($n = 30$) regarding the eight periods of subjective fatigue ratings with the Kendall coefficient of concordance (W). The probability of significance associated with an observed W was estimated as chi-squared.

We estimated the relationships among the workers' characteristics and subjective fatigue ratings by computing Spearman rank correlation coefficients. The subjective rating value used in computing these correlation coefficients was the median of all eight period ratings given by the worker. Worker age, years and months at present job, years and months since starting a job similar to the present job, and total number of years of employment similar to the present job were examined.

Aircrew Members. Fatigue ratings were collected from 28 male USAF aircrew members. They were pilots, copilots, and navigators who flew more than 37 usable sorties during eight formation airdrop missions in the C-130H Hercules tactical transport. Pairs

(before flight, after flight) of fatigue ratings were collected from the aircrew members. These ratings were separated into 25 morning pairs and 11 afternoon pairs. The participation by a crewmember in more than one morning or afternoon formation mission was represented as a single pair of ratings by taking medians of the multiple ratings. Thus, each crewmember was represented only once in the morning set or once in the afternoon set. Nine crewmembers were common to both sets.

Before-flight ratings were collected during pre-flight briefings which began between 0330 and 0645 (morning missions) or between 1200 and 1610 (afternoon missions). After-flight ratings were collected during post-flight debriefings which began between 1130 and 1430 (morning missions) or between 1930 and 0100 (afternoon missions).

Median fatigue ratings were calculated for before- and after-flight reports for both morning and afternoon sorties. These before- and after-flight medians were compared using the Wilcoxon matched-pairs, signed-ranks test. Median morning and median afternoon values for the nine crewmwmbers who provided ratings for both periods were calculated from the before- and after-flight values. These morning and afternoon medians were compared using the Wilcoxon test. Correlations between before- and after-flight ratings were estimated as the Spearman rank correlation coefficient. The intercorrelation between the checklists was also calculated using this coefficient.

Combined Data. A single median value for each of the 28 aircrew members was calculated from the before- and after-flight pair of ratings (two pairs for nine of the crewmembers). Similarly, a single median value for each of the 30 office worker participants was calculated from the eight ratings given. These two sets of ratings were compared using the Mann-Whitney U-test (Dixon, 1981; Siegel, 1956).

RESULTS

Office Workers. Subject age was normally distributed, ranging from 18 to 61 years, with a mean of 34.1 years. They had been employed at their present jobs for an average of 4.8 years, ranging from 2 months to 21 years. Subjects averaged 7.2 years since they first started jobs similar to their present jobs, with a range of zero years to 24.4 years. Their mean length of employment at their present jobs and similar ones in the past was 8.9 years, with a range of 3 months to 27.8 years.

The results of the Friedman tests indicated a significant difference within each day for each of the checklists. The p values for the day 1 and day 2 Friedman analyses were less than .001, except 20PT, day 2, which was .008. Overall, the trends in reported fatigue for the checklists were similar within days 1 and 2. Subjective fatigue appeared to follow a similar pattern for both the 7PT and 20PT checklists in that subjects reported gradually increasing fatigue throughout the day. A visual in-

spection of the data suggested greater levels of fatigue on day 2 for the 7PT, but the difference was not significant ($p = .43$).

The intercorrelation between the 7PT and 20PT was relatively high ($r = -.80$), while the correlations between days 1 and 2 were relatively low (7PT, $r = .34$; 20PT, $r = .33$). Inter-rater agreements (Kendall W) with regard to the eight periods of subjective fatigue ratings were statistically significant for each checklist, but very low. The examination of the relationships among the subjects' characteristics and their fatigue ratings indicated no useful patterns.

Aircrew Members. The results of the Wilcoxon tests indicated a significant difference from before to after flight for each checklist for both the morning and afternoon flights. The 2-tail p values for the respective Wilcoxon tests were:

	p-Values	
Scales	Morning (n = 26)	Afternoon (n = 11)
7PT	.008	.003
20PT	< .001	.004

There appeared to be an effect of time of day on the aircrew fatigue ratings. The median levels of fatigue reported for morning and afternoon missions differed for the nine aircrew members providing data from both periods. The morning missions were characterized by greater fatigue than the afternoon meetings (due, perhaps, to early morning sleep disruption):

	Ratings		2-tail
Scales	a.m.	p.m.	p Value
7PT	5.0	3.5	.036
20PT	7.0	11.25	.012

The correlation between the two checklists ($r = -.79$) was about the same as for office workers. The correlations between before- and after-flight ratings were relatively high (7PT, $r = .48$; 20PT, $r = .66$)

Combined Data. The comparisons of single median ratings given by office workers to those given by aircrew members revealed a consistent pattern. The aircrew members reported more fatigue:

Scale	Office Workers (n = 30)	Aircrew Members (n = 28)	2-tail p Value
7PT	3.0	4.0	< .001
20PT	12.5	9.6	< .001

DISCUSSION

The checklists used in this investigation were sensitive to variations in office worker fatigue throughout the workday. They were also sensitive to variations in aircrew fatigue before and after flights and for the comparison of morning to afternoon flights. The p values for Friedman tests and Wilcoxon tests indicated highly reliable differences for the respective comparisons. The findings that aircrew members reported greater fatigue than office workers lent face validity to the conclusion that the checklists were sensitive to fatigue. However, this result was confounded by gender effects (twenty-nine of the 30 office workers were female, and all of the aircrew members were male). In general, the results of the investigation suggested that both checklists are equally sensitive and reliable.

Dixon, W.J. (ed.) (1981). BMDP Statistical Software, 1981. Berkeley, California: University of California Press.

Freidmann, G., Globus, G., Huntley, A., Mullaney, D., Naitoh, P., and Johnson, L. (1977). Performance and mood during and after gradual sleep reduction. Psychophysiology 14, 245-50.

Martin, I., and Venables, P.H. (1980). Techniques in Psychophysiology. Chichester: John Wiley and Sons.

Pearson, R., and Byars, G.E., Jr. (1956). The Development and Validation of a Checklist for Measuring Subjective Fatigue. Randolph AFB, Texas: USAF School of Aerospace Maedicine, Technical Report 56-115.

Siegel, S. (1956). Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill.

Storm, W.F. (1983). Aircrew fatigue during extended transport, tactical, and command post operations. Neuilly sur Seine, France: NATO Advisory Group for Aerospace Research and Development, Conference Proceedings No. 338.

Storm, W.F., and Merrifield, J.T. (1980). Fatigue and Workload in Four-man Cockpit Crews (Volant Galaxy). Brooks AFB, Texas: USAF School of Aerospace Medicine, Technical Report 80-23.

A Descriptive Model for Time to Switch Among Subtasks[1]

Kevin Boettcher
Honeywell Systems and Research Center
MN65-2400, 2600 Ridgway Parkway, Minneapolis, MN 55413

ABSTRACT

The additional mental load induced when mental switching is required is investigated. Data are presented that indicate a significant contribution to response time due to switching overhead. A descriptive model, derived from the information theoretic characterization of systems, is suggested and validated.

I. Introduction

Complex human information processing tasks often require attention switching and mental re-orientation. In command and control structures, for example, information and messages can arrive from multiple sources such that dissimilar processing requirements are interspersed. The process of re-orientation requires mental resources and contributes to workload. To the extent that this contribution is significant, it must be taken into account when assessing whether task demands exceed processing limits. This paper presents experimental evidence, using a simple test situation, that switching overhead can be significant. A descriptive model is then tested and validated that reflects the additional time required for switching. The model is information theoretic in origin and relates switching frequency to processing time through a specific functional form. Because the model is not particular to the task considered, it can be applied more generally to situations where a human is required to alternate among several distinct subtasks.

II. Experimental Task with Switching

To investigate the significance of switching, the experimental task shown in Figure 1 was used. On a given trial one of three possible "thresholds" was

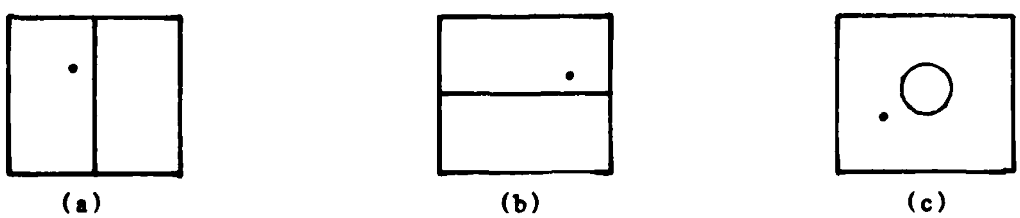


Figure 1 Switching Task

presented to a subject (within a fixed, square border): (a) a vertical line (b) a horizontal line or (c) a circle. For each situation a dot was also displayed, and the subject was to select one of two possible responses according to the position of the dot relative to the threshold: (a) left or right (b) up or down, or (c) inside or outside. Responses were registered by depressing one of two mechanical, horizontally-arranged buttons. Dot position was deter-

[1]This work was supported by the Office of Naval Research under grants ONR/N00014-77-C-0532 (NR 041-519) and ONR/N00014-84-K-0519 (NR649-003).

mined randomly, and independent of threshold type. The relative mixture of threshold types was controlled by the experimenter as the independent variable. Because of the arrangement of the response mechanism vis a vis the three threshold types, re-orientation by the subject was required each time the threshold type changed; switching among subtasks requires that the subject recall the proper association of response buttons to threshold type before completing the processing required by the specific dot position presented.

Experimental conditions were established according to the relative mixture of threshold types. An experimental run consisted of a pre-announced number of trials (usually 100 or 200) at a given condition. Prior to a run, subjects were told only which threshold types would be included, but not their relative frequency. On each trial the threshold and dot appeared simultaneously, and the subject was instructed to respond as quickly as possible, but with certainty. An 800 ms blanking interval was used between trials. (A complete description of experimental procedures and results is given in Boettcher and Tenney (1985).)

III. Binary Switching

With three threshold types, there are three possible combinations involving two types each. After a suitable training period, each combination was investigated by measuring response time for five mixtures at each binary combination. Average response time observed for each condition tested is shown in Figure 2 for one subject, RF. The results shown in the figure, which

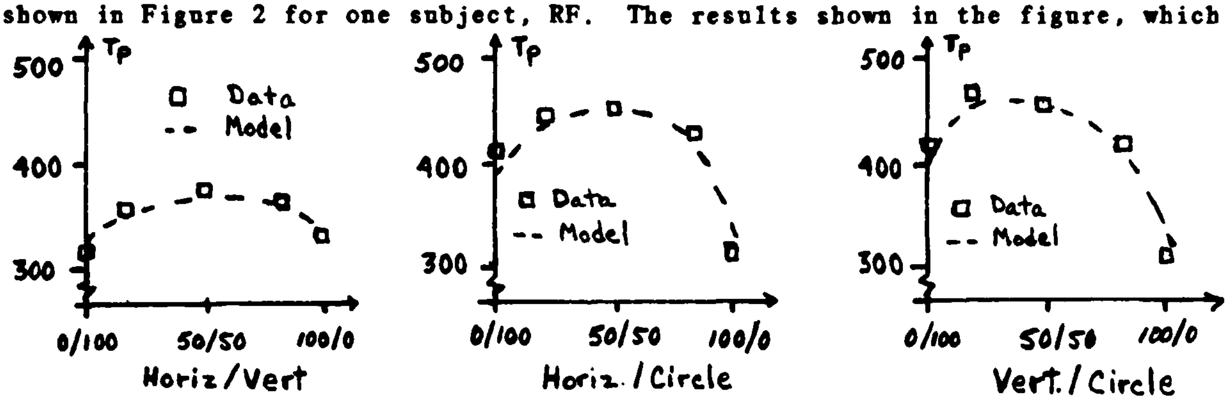


Figure 2 Binary Switching Results

are typical of those observed for other subjects, indicate that an overhead in processing time due to switching is present. Furthermore, it is apparent that this overhead can be significant. For the cases where the circle threshold was used, each trial in a 50/50 mixture requires roughly 25% more time to process, on the average, beyond that which is accounted for by simply averaging the time required for individual threshold types.

Descriptive Model

Instances that involve switching have been studied by other investigators (e.g. Moray (1969), Remington (1969)) and results similar to those shown in Figure 2 have been observed previously. Building on this earlier work, and using the specific data presented above, a novel descriptive model for the switching phenomenon is now presented that is relatively simple in form, yet is grounded in concepts derived from the analysis of complex systems.

When described in information theoretic terms, systems with switching exhibit an increase in processing activity that is characterized by the Shannon entropy function (Boettcher and Levis (1982)). Denoted by H, this function represents the uncertainty present in a random variable that can take two values, one with probability p and the other with probability 1-p:

$$H(p) = -p \cdot \log_2(p) - (1-p) \cdot \log_2(1-p) \tag{1}$$

Interpreted in the context of switching overhead, p designates the relative frequency of a given sub-task's execution and H characterizes the additional processing activity required to switch to that sub-task. If a sub-task is executed exclusively (p=1 for that sub-task), no re-orientation is necessary and hence no switching is present (H = 0). Similarly, if a subtask is never executed (p=0), there is also no re-orientation necessary. Finally, if a task is such that some sub-task is executed 50% of the time, the additional overhead for switching is at a maximum for that sub-task (H = 1).

Using the functional form in eq.(1), and assuming that use of mental processing resources is directly related to observed processing time, a model for switching overhead can be formulated. Specifically, for the present task a model for the binary combination of horizontal and vertical thresholds is

$$T_p = p_v\mu_v + p_h\mu_h + \alpha_v \cdot H(p_v) + \alpha_h \cdot H(p_h) \tag{2}$$

where v and h designate vertical and horizontal thresholds, p_i is the fraction of threshold i's use, and μ_i is the average time required when threshold i is exclusively presented. T_p is the average response time for each trial under the condition delineated by p_i, and the contribution to this average due to switching is given by the last two terms. Note that while the variation in switching overhead is governed by H, the parameters α_i account for the level of this contribution.

Applying the model form in eq.(2) to the switching results shown in Figure 2, three parameters (two values of μ and a sum of two α values) were estimated from the five experimental conditions examined for each binary combination. (Because of the symmetry of H, it is not possible to distinguish α values in binary mixture data. For example, in eq.(2) only $\alpha_v + \alpha_h$ can be estimated.) The identified models for each binary combination have been superimposed on the data in Figure 2. It is evident that the model form of eq.(2) represents a reasonable fit to the data observed.

IV. Three-Way Switching

The switching model represented by eq.(2) extends readily to situations involving more than two sub-tasks. For example, the general three-way model as it applies to the experimental task at hand is given by

$$T_p = p_v\mu_v + p_h\mu_h + p_c\mu_c + \alpha_v\, H(p_v) + \alpha_h\, H(p_h) + \alpha_c\, H(p_c) \tag{3}$$

A test of the model in eq.(3) was conducted as follows. Figure 3 illustrates graphically the simplex of possible three-way mixtures. For a given point in the simplex, the corresponding p_h, p_v, and p_c values are obtained by measuring the normalized distance of that point to each side. For example, point 3

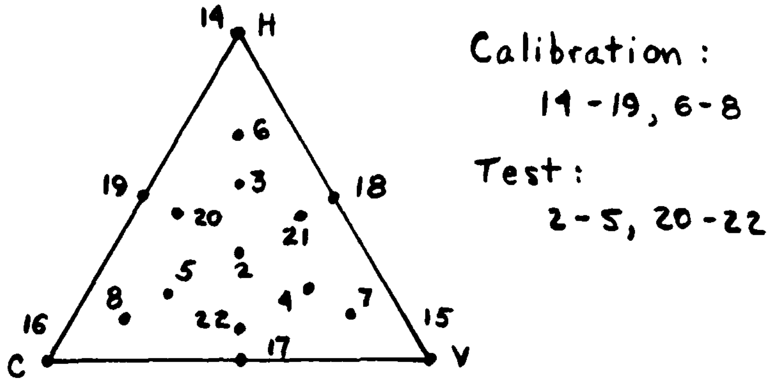


Figure 3 Simplex of Three-Way Mixtures

corresponds to p_h, p_v, and p_c values of 0.5, 0.25, and 0.25, respectively.

In a single session, subjects were given a set of mixtures chosen from various points in the simplex. The results from some conditions in the set were used to estimate model parameters, and the identified model was then used to predict the results for the remaining conditions in the set. The conditions selected for model parameter estimation were chosen to be deliberately distinct in the simplex from those used for model testing. Test conditions tend toward the "middle" of the simplex, while the calibration conditions are positioned near the "edges," as indicated in the figure. This was done in order to examine the ability of the model to predict behavior outside the region from which it was constructed. Subjects were not aware which conditions were for model calibration and which for test. Furthermore, the calibration and test conditions were interspersed with each other.

Estimates for model parameters μ were obtained directly from observed data; that is, the mean value of processing time observed was taken as the value of μ. Estimates for α values were obtained using a least squares method. The resulting models are tabulated by subject and session in Table 1,

Table 1 Model Parameter Estimates for Three-Way Switching

Session	Subject	$\hat{\mu}_h$	$\hat{\mu}_v$	$\hat{\mu}_c$	$\hat{\alpha}_h$	$\hat{\alpha}_v$	$\hat{\alpha}_c$	95% Conf for α
1	BM	289	246	340	3.9	32.3	52	$\pm$ 18
2	BM	273	257	320	16.4	29.8	19.6	$\pm$ 18
3	PP	309	294	370	29.4	24.5	46.9	$\pm$ 15
4	PP	284	272	350	3.4	23.1	40.7	$\pm$ 19
5	RF	310	310	357	37.8	14.3	29.3	$\pm$ 7

along with simultaneous 95% confidence intervals on the estimates of α values.

Using the identified models for each respective session, predicted values of processing time were obtained for other conditions, along with 95% confidence intervals. These predictions, denoted T_p, are shown in Table 2 together with the actual average processing times observed (T_p). Comparing the predicted and observed processing time values of Table 2, there is substantial agreement between model and data. More than 75% of the tests have T_p within the 95% confidence bounds; in many cases, values are much closer. There does not seem to be any condition where a systematic violation occurs.

Table 2 Switching Model Predictions and Test Results

Sess	Subj	Cond	T_p	$\hat{T}_p$	95%	Sess	Subj	Cond	T_p	$\hat{T}_p$	95%
1	BM	20	382	378	±13	3	PP	20	424	423	± 11
		21	330	339	"			21	377	388	"
		22	355	378	"			22	406	415	"
		2	370	370	±12			2	450	415	± 10
2	BM	21	315	328	±16	4		20	373	368	± 8
		22	353	344	"			21	352	334	"
		2	330	342	±13			22	372	374	"
		3	357	337	"			2	388	362	± 9
						5	RF	5	399	405	± 10
								2	390	399	"

This lends support to the model form as a simple characterization of the additional time required for mental switching among dissimilar subtasks.

V. Summary

This paper has considered the overhead required in mental processing resources to switch among dissimilar subtasks. A specific experimental task has been used to document that such switching can be significant and a descriptive model for this switching, based on information theoretic concepts has been suggested. Tests of the model indicate support for its validity and applicability. Of particular note is that the integrity of sub-task execution is preserved and that variation in switching overhead is related only to the relative frequency with which the sub-task is executed. Thus it may be possible to construct models for complex tasks that involve switching by simply augmenting existing models of component subtask with the appropriate switching terms. By accounting for any switching phenomena, a more accurate assessment of task workload can be made.

References

Boettcher, K.L. and A.H. Levis, 1982. Modeling the interacting decisionmaker with bounded rationality, IEEE Trans on Sys, Man and Cyber, 12:334-344.

Boettcher,K.L. and R.R. Tenney, 1985. "A Procedure-Based Approach to Human Information Processing Models," Lab for Infor. and Dec. Sys., MIT, LIDS-P-1499.

Moray, N. 1969. Attention: Selection Processes in Vision and Hearing, London: Hutchinson Educational.

Remington, R.J., 1969. Analysis of sequential effects in choice reaction time. J. Exp. Psychol., 82:250-247.

Effects of stimulus quality and priority manipulations in dual-task performance.

Andrew J Tattersall
Donald E Broadbent
David B Porter

University of Oxford

Two experiments were carried out in order to investigate the effects of two types of stimulus degradation on interference between two visually presented recall tasks. Resultant Performance Operating Characteristics (POCs), plotted for varying priority conditions between the tasks, showed a trade-off between the two tasks, indicating that the tasks were sharing some common resource. However, neither type of degradation interacted with this priority trade-off, suggesting that stimulus degradation affects some resource which is not competed for by the non-degraded task. These results are interpreted to support a multiple resources concept of information processing.

INTRODUCTION

In many complex systems, such as aircraft or industrial control rooms, operators are often required to perform two or more tasks at the same time. This requirement to combine tasks may result in interference between the tasks by placing excessive demands on limited processing resources which may be shared between them.

There is a clear need, therefore, to investigate the manner of this interference between various concurrently performed tasks and to develop guidelines based on theoretically derived principles which would aid both system designers and operators.

A promising approach to the study of task combination is that proposed by Norman and Bobrow (1975, 1976). This is the method of applying varying priority manipulations to the two tasks in order to generate Performance Operating Characteristics (POCs). A POC is a curve which depicts the change in the level of performance on one task when the level of performance on another, simultaneous, task is changed. It is argued (e.g. Navon and Gopher, 1979) that conclusions may be drawn from the shape of these resulting curves about the extent to which two tasks use common resources. For example, difficulty manipulations that affect a common resource are predicted to interact with the priority manipulation and produce a change in slope of the POC, whereas manipulations affecting resources which are only relevant to one task should not affect the slope of the POC, but may cause a shift in the curve by a constant amount.

A number of studies have investigated input and output interference (i.e. sensory and motor overlap) in various dual-task situations (see Wickens, 1980 for a review), however, little empirical work has been reported which involves interference in central processes.

Fitzgerald, Tattersall and Broadbent (in preparation) report a number of experiments which applied the POC technique in order to identify separate central processing mechanisms in temporary memory. In these experiments, two lists of items were presented for later recall, with varying priorities for each sub-list. In each experiment a trade-off

between the two tasks was found, indicating that at some point the tasks were using some common mechanism. When the lists were presented simultaneously, effects of recall order interacted with priority, so that the high priority items were more harmed by recall order than were the low priority items. With successive presentation of items, that is, one sub-list presented before the other, priority was not found to interact with recall order. However, order of presentation did interact in this way, with the second presented items showing a larger effect of recall order than the first presented. They argue that the items are held in three types of temporary storage, an articulatory loop that is not disturbed by output (Fitzgerald and Broadbent, 1985), a form of temporary memory also not disturbed by output, and an input-output buffer that holds items coming into the system until there is fresh input or output. If there is successive presentation of items then the buffer is dedicated to the most recent items, and material passes from it to the two other forms of memory, where its maintenance may be influenced by sub-list priority. When material is to be reported, then for the first reported items, any item in any of the kinds of memory can be retrieved. However, items passing through the buffer, either input or output, produce a loss of items being held there. Therefore, output order affects primarily the second of the two successive sub-lists.

If the sub-lists are presented simultaneously, however, then the input-output buffer, as well as the other two forms of storage, is partitioned between sub-lists according to their priority. Therefore, output order will affect the high priority sub-list more seriously since there are more items from that sub-list in the buffer.

This suggests that recall order may affect shared resources which are involved in encoding during presentation, but does not affect shared resources which are involved in retention of encoded material. The present experiments aim to examine further the role of the input-output buffer using the POC technique. If degradation interacts with the priority and recall order interaction, then resources involved in maintaining items in the buffer may also be involved in encoding visual material. If, however, this difficulty manipulation does not alter the trade-off, then it is likely that it is affecting some other part of the system which is more specific to one task than the other.

Two different types of degradation were used in the present experiments, either pattern degradation, which was the addition of random dots around the stimulus, or contrast reduction. There are suggestions in the literature (e.g. Hardzinski and Pachella, 1980; Pashler, 1984) that visual noise masking may have different effects (i.e. on different substages of encoding) to contrast reduction. It is possible that these differential effects may be observed if visual noise degradation requires resources which are shared with the input-output buffer and which are not affected by reduced contrast.

EXPERIMENTAL DETAILS

The basic procedure was the same for both experiments, the main difference between them was the type of degradation used.

In Experiment 1, 12 female members of the Oxford Subject Panel served

as paid subjects. Their ages ranged from 20 to 42 years with a mean age of 30 years. Twelve different subjects, also female members of the Oxford Subject Panel, took part in Experiment 2. Their ages ranged from 18 to 38 years with a mean age of 26 years. All subjects reported good normal or corrected vision, and none had taken part in previous experiments of this type.

A Sinclair Spectrum was used for controlling stimulus presentation in Experiment 1. A BBC microcomputer was used in Experiment 2. In order for comparisons to be made, user definable graphics were used to create characters which were identical to the Sinclair Spectrum character set. The stimuli were displayed on a Sony, 12 inch, black and white monitor.

The stimulus material was taken from a character set of 20 letters (the letters of the alphabet minus I,O,Q,S,U,Z) and the digits 1 to 9. 108 lists were constructed, each containing 4 letters and 4 digits. The letters and digits were assigned randomly to the lists with the constraint that no item appeared more than once in any list. The lists were presented in a different randomized order to each subject. They were presented as black (or shades of grey) on white in the centre of the screen with the letters slightly above the digits. The characters subtended 15 minutes of arc at the eye horizontally, 17 mins vertically, with a vertical separation between the characters of 24 mins, all with a viewing distance of one metre.

In Experiment 1, the characters were degraded by adding 6 extra pixels to the 8*8 dot matrix of the character. These pixels were chosen at random for each list. They were drawn in black unless the co-ordinate coincided with the character, in which case the character colour (black) was inverted to the background colour (white).

Experiment 2 involved degradation of one set of stimuli by reduction of the contrast of that set. This was achieved by using the BBC microcomputer's colour set to give different shades of grey on the black and white screen. The normal stimulus set used black characters, the degraded set used the yellow colour to give a light grey colour on the white background.

In both experiments one set of stimuli (either letters or digits) was degraded, and the other set was non-degraded. Half the subjects in each experiment received the stimuli with letters degraded, the remaining subjects with digits degraded.

The first 12 lists constituted a single-task condition. The subjects received the remaining lists in three 'priority' conditions, in which the letters and digits had different points attached to them. These were: (i) letters 9 points and digits 3 points each, (ii) equally valuable with 6 points each, and (iii) digits 9 points and letters 3 points. Each condition consisted of 32 lists, 8 practice and 24 test lists. Half the subjects received letters high priority first, the other half with digits high priority first.

Recall order (i.e. whether letters or digits were to be reported first) was determined by a post-list cue: "<---" represented letters first, and "--->" digits first. Recall cue order was randomized but with half the trials in each condition of each type.

The subjects were tested individually. The basic task required them to remember the letters and digits in each list and to write them down at

the end of each list presentation in the order required. Each list began with a warning tone, a blank screen for 500 msec, then the stimulus list. This consisted of 4 letter-digit pairs, with each letter and digit presented simultaneously for 1000 msec, with an inter-stimulus interval of 20 msec. The list was followed by a recall cue. Presentation of the lists was self-paced. Responses were scored after each list, that is, subjects received feedback about their performance and the priority values. Subjects were given a short break and new priority instructions before each experimental condition.

RESULTS

Analyses of variance were performed for each experiment, in which between-subject factors were direction (whether letters or digits were initially more valuable) and type of item degraded; within-subject factors were priority condition ("9:3","6:6" or "3:9"), order of report (first- or second-reported), type (letters or digits) and serial position.

There was a clear effect of priority in both experiments, in Experiment 1 (pattern degradation), $F=12.55$, $df=2,16$, $p<0.001$; in Experiment 2 (contrast degradation), $F=24.10$, $df=2,16$, $p<0.001$. This trade-off indicates that letter and digit recall are not independent and that some common resource is involved.

Both experiments showed a significant interaction between priority and recall order, Expt 1: $F=5.58$, $df=2,16$, $p<0.05$; Expt 2: $F=3.97$, $df=2,16$, $p<0.05$; replicating the results of Fitzgerald et al, that is, for simultaneous presentation of items, output order affects recall of high priority items more than recall of low priority items.

However, neither experiment showed a significant interaction between type of item degraded and priority, Expt 1: $F=1.74$, $df=2,16$, $p>0.1$; Expt 2: $F=0.02$, $df=2,16$, $p>0.1$; or a significant interaction between type of item degraded, priority and recall order, Expt 1: $F=0.85$, $df=2,16$, $p>0.1$; Expt 2: $F=0.83$, $df=2,16$, $p>0.1$. These results suggest that degradation is affecting some resource which is not competed for by the non-degraded task.

Other, though less important, significant effects included recall order, for which recall of first-reported items was superior to recall of second-reported items (Expt 1: $F=52.66$, $df=1,8$, $p<0.001$; Expt 2: $F=63.74$, $df=1,8$, $p<0.001$); and serial position (Expt 1: $F=7.94$, $df=3,24$, $p<0.001$; Expt 2: $F=6.33$, $df=3,24$, $p<0.01$).

DISCUSSION

These experiments confirm that for simultaneously presented lists of items, recall of the high priority list is more harmed by prior recall of the low priority list than the low priority list is by prior recall of the high priority list.

Secondly, the results suggest that the degradation does not affect those resources which are also used for maintaining recently input and about-to-be output items. As priorities are changed, an increase in performance on the degraded material requires no greater loss for the other material than would an increase in performance on non-degraded

material. Thus, the shared resources, as indicated by the priority and recall order interaction, applied to the degraded material, come from some mechanism which is not itself affected by degradation.

Thirdly, no evidence has been found to support the notion that pattern degradation and contrast reduction are affecting different stages of encoding.

REFERENCES

Fitzgerald,P., & Broadbent,D.E. (1985). Order of report and the structure of temporary memory. Journal of Experimental Psychology: Learning, Memory and Cognition, 11, 217-228.

Fitzgerald,P., Tattersall,A., & Broadbent,D.E. Separating central mechanisms by POCs: Evidence for an input-output buffer. To be submitted.

Hardzinski,M., & Pachella,R.G. (1980). The manipulation of stimulus quality and the definition of stimulus encoding operations in memory scanning experiments. Perception and Psychophysics, 27, 232-240.

Navon,D., & Gopher,D. (1979). On the economy of the human-processing system. Psychological Review, 86, 214-255.

Norman,D.A., & Bobrow,D.G. (1975). On data-limited and resource-limited processes. Cognitive Psychology, 7, 44-64.

Norman,D.A., & Bobrow,D.G. (1976). On the analysis of performance operating characteristics. Psychological Review, 83, 508-510.

Pashler,H. (1984). Evidence against late selection: Stimulus quality effects in previewed displays. Journal of Experimental Psychology: Human Perception and Performance, 10, 429-448.

Wickens,C.D. (1980). The structure of attentional resources. In R.Nickerson (Ed.), Attention and Performance VIII. Hillsdale,N.J.: Erlbaum.

The Effect of Task Difficulty On The Steady State Visual Evoked Response

Andrew M. Junker
AF Aerospace Medical Research Laboratory WPAFB

Kevin M. Kenner
Synergy Inc., Dayton Oh.

Elizabeth J. Casey
Univ. of Illinois

ABSTRACT

Steady State Visual Evoked Responses to sum-of-sine wave modulated light were measured for 10 subjects as they performed a decision making task. Describing function phase values demonstrated a positive shift under task conditions. The gain of the frequency response exhibited attenuation with increasing task difficulty for half the subjects.

INTRODUCTION

The Air Force has placed an emphasis on developing performance assessment measures. Workload metrics developed to date have consisted of either subjective measures such as the Subjective Workload Assessment Test (Reid et. al.,1981) or objective, but obtrusive, measures that require an active focus of attention by the subject, like the Critical Task Set (Shingledecker et. al.,1983). An unobtrusive physiological measure is needed. It is possible to detect a response in the human electroencephlogram (EEG) to evoking stimuli. When the evoking stimulus is sinusoidally modulated light, the result is called a steady state Visual Evoked Response (VER). Earlier work has shown the value of the steady state VER as a potential workload metric (Kenner, Junker and Levison, 1985, Junker and Peio, 1984, Wilson and O'Donnell, 1981). The objective of this study was to determine whether the steady state VER to sinusoidally modulated light varied systematically with difficulty of a decision-making task.

METHOD

A stimulus presentation device simultaneously delivered the evoking stimulus (flickering lights) and a video task display (Figure 1). This presentation was achieved by combining the two images via an 18 cm x 26 cm half-silvered mirror at 45 degrees to the two images. The evoking stimulus was two 26 cm fluorescent light tubes hung horizontally and mounted 4 cm behind a 25 cm x 27cm translucent, diffusing screen; which

distributed the light, as evenly as possible, over the visual field. The intensity of the lights (average intensity 40 ft-Lamberts) was measured by a United Detector, model PIN 10D, high speed photo cell, placed at the subject's viewing point. The average intensity of the evoking lights was sufficiently low that a subject could still comfortably discern the video task-display within the same visual field. The video decision-making task was displayed on an Audiometrix 11 in x 11 in video monitor.

Ten subjects participated in one or more two-hour data-collecting sessions. Subjects were seated in a darkened test chamber facing a 15 x 15 cm window. Behind the window was the stimulus presentation device. For the Lights Only (LO) condition subjects were instructed to "relax and fixate on the center of the screen." For the Decision Making (DM) conditions the subjects were instructed to concentrate on the tasks. At the end of each ninety-second trial the subject's score appeared on the screen.

Beckman silver/silver chloride electrodes were used with a Grass model P511 AC amplifier (amplification x50,00). The EEG was recorded with electrodes at Oz with right mastoid as reference and left mastoid as ground (less than 5k ohm electrode resistance). The Sum-Of-Sines (SOS) wave was generated, and data collected, on a PDP 11/60 computer. Signals from the 11/60 were filtered (cut off at 40 Hz) and then fed into a Scientific Prototype, model GB, tachistiscope/light driver, which was modified so the average intensity and depth of modulation could be adjusted. The two channels of data (photo cell and EEG) were filtered (cutoff at 25 Hz), then digitized and stored for analysis on the PDP 11/60. The collected data was fast Fourier transformed, ensemble averaged and plotted using a PDP 11/34 computer. The Dynamic Decision Making task was generated on the PDP 11/60 computer.

The sinusoidally modulated light served as the evoking stimulus. The lights were modulated using a SOS wave composed of 10 harmonically non-related frequencies. All 10 of the frequencies were multiples of the fundamental frequency of 0.0244 Hz. The component frequencies ranged from 6.25 to 21.75 Hz, with intermediate frequencies at 7.75, 9.50, 11.50, 13.25, 14.75, 16.50, 18.25, and 20.25 Hz. None of these component frequencies contained a sum or difference of any of the other component frequencies; this restriction on sine wave selection was implemented to avoid first order nonlinearity interactions of the spectral power.

For every data collecting trial starting phase values for each of the 10 component sine waves were randomized with a uniform random number generator, insuring that the time sequence of the flickering light presentation was random from trial to trail. By utilizing randomized phase with the summing of the 10 sinusoids a peak depth of modulation of 13 %, per sinusoid was achieved.

A Supervisory Control task was used as the decision making task (Pattipati et. al., 1979). This task involved the problem

of allocating attention among multiple tasks in a supervisory control system. Two levels of difficulty were used. In the "easy" condition it was possible to successfully allocate attention among the multiple tasks. In the "hard" condition the time required exceeded the time available and it was not possible to complete all allocations successfully.

Manipulations of the fast Fourier transforms of the evoked response potential signals and photo cell signals provided describing functions which are a complex measure of the input-output relationship of the visual-cortical system. The focus of this project was on the amplitude ratio and the phase angle measures obtained from these computations. Estimates of mean values for the gain and phase computations across replications were computed. For indication of mean variability standard errors were computed. The amplitude ratios indicate gain sensitivity of the system at the component frequencies. The phase angles relate to neurophysiological dynamics and transmission latency between stimulus and VER measurement.

RESULTS AND CONCLUSIONS

Figures 2 and 3 show the visually evoked, steady state responses in the frequency domain of two subjects to the three experimental conditions. The data in each graph represents the average of eight trials from sessions in which responses to flickering lights were strongest. Vertical lines at each data point represent standard error about the mean. Absence of error bars indicate variability less than the symbol height.

The data for one subject (Figure 2) show the general pattern we expected. Over most of the range of frequencies, the gain for the Lights Only (LO) condition was highest and the gain of the response to the lights during the more difficult DM condition was lowest. The phase of the response was equivalent for the two DM conditions while it shifted negatively in the LO condition.

These results indicate an attenuation in response to flickering lights as attention is drawn from the lights to a task. The attenuation strengthens as task difficulty is increased. In addition, the frequency response of the visual cortical system is increased (less negative phase shift) when attention to a task is required although task difficulty did not seem to affect the phase.

The graph of Figure 3 demonstrates that the simplicity of these results was not consistent in all subjects. In general, the gain of the LO response was higher than the responses of the DM conditions at most frequencies. However, there were several cases in which the power of the LO response was lower at certain frequencies. Between subjects there was little consistency in the gain of the DM response patterns.

The most consistent pattern for 9 of the 10 subjects was the phase shift between the LO condition and the DM conditions. The phases of the two decision making tasks, however, were equivalent for all subjects.

The results of this study do not provide conclusive evidence of the usefulness of steady state VER's as a measure of workload. However, they do suggest that this line of research may eventually produce a useful measure. The data considered in this study represents average responses over 90 second periods. Average changes were detected, however, averaging has the tendency to reduce extreme variations. Task demands changed dynamically over the 90 second periods. In the next set of studies, frequency response measures at selected frequencies over time locked task demand periods, using Lock-in amplifiers, will be investigated. The data from the present study will guide the selection of frequencies to be used for each subject in the planned investigations.

REFERENCES

Junker, A.M. and K.J. Peio 1984, "In Search of a Visual-Cortical Describing Function, A Summary of Work In Progress", Proc. of 20th Ann. Conf. on Manual Control, NASA Conf. Publ. 2341, 1984.

Kenner, K.M., Junker, A.M., and W.H. Levison 1985, "A Linear Dynamic Model for the Visual-Cortical Evoked Response System", NACON Proceedings 20-24 May 1985, Dayton Ohio.

Pattipati, K.R., Kleinman, D.L., and A.R. Ephrath 1982, A Dynamic Decision Model of Human Task Selection Performance", Univ. of Conn. Tech. Report, Dept. of EECS, TR-82-1, March 82.

Reid, G.B., Shingledecker, C.A., and Eggemeier, F.T. 1981, "Application of Conjoint Measurement to Workload Scale Development", Proc. of 1981 Hum. Factors Soc. Ann. Meeting, Oct. 1981(a), pp.522-526.

Shingledecker, C.A., Acton, W.H., and Crabtree, M.S. 1983, "Development and Application of a Critereon Task Set for Workload Metric Evaluation", Proc. of 2nd Ann. Aerospace Behavioral Engineering Tech. Conf., Long Beack, Ca., 1983.

Wilson, G.F., and O'Donnell, R.D. 1981, "Human Sensitivity to High Frequency Sine Wave and Pulsed Light Stimulation as Measured by the Steady State Cortical Evoked Response", Tech. Report AFAMRL-TR-80-133, 1981.

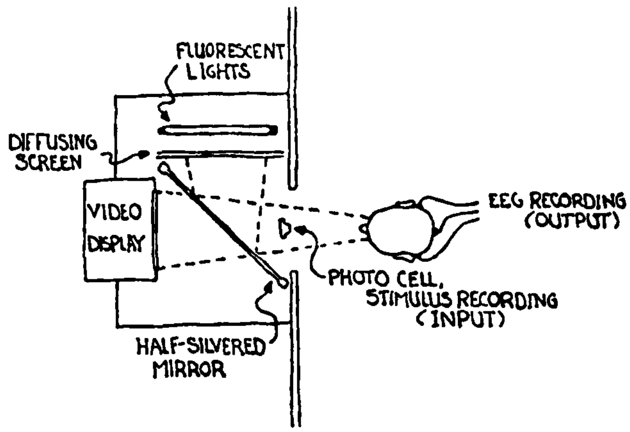


Figure 1. Experimental Setup

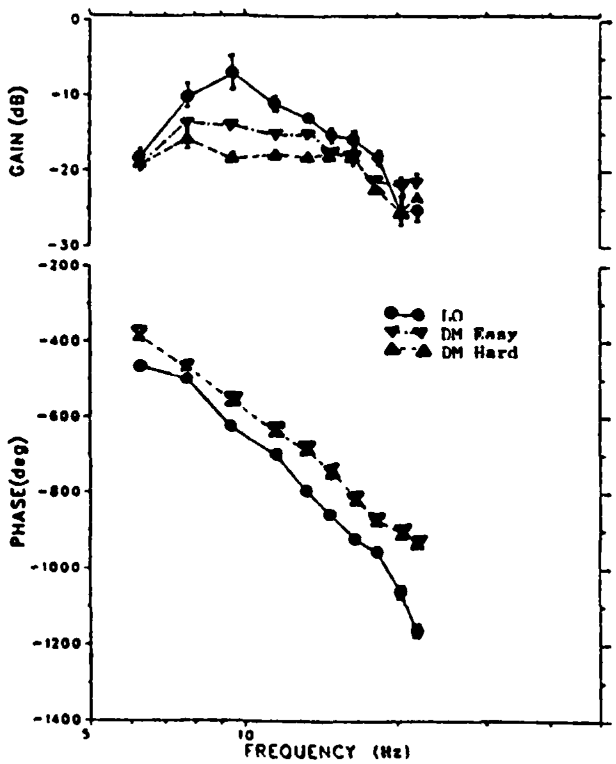


Figure 2.

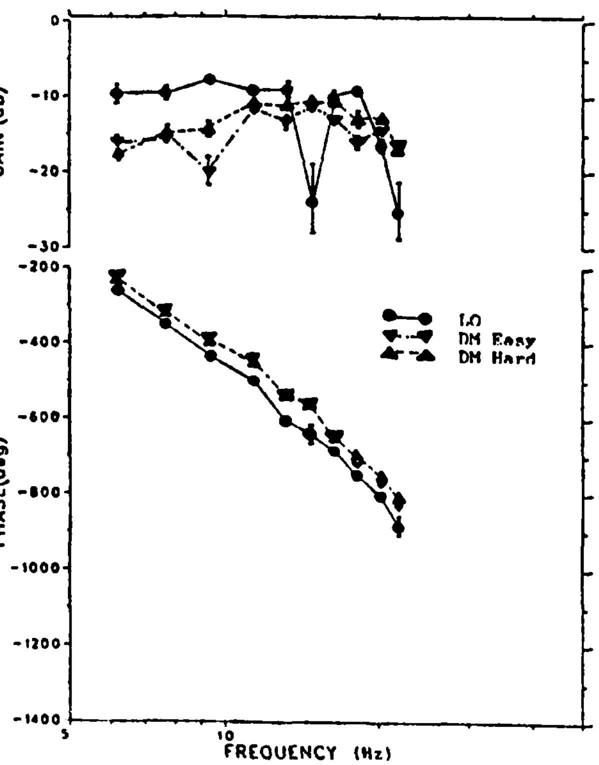


Figure 3.

VER DESCRIBING FUNCTIONS.
Lights Only (LO)-circles, Decision Making (DM)-triangles

Behavioral Issues Associated with Isolation and Confinement: Panel Discussion

Jack Stuster, Ph.D.
Anacapa Sciences, Inc.

Purpose:

The purpose of our discussion is to stimulate scientific and DOD interest in behavioral issues concerning military environments characterized by conditions of isolation and confinement. Though our interests include the full range of behavioral issues, our discussion focuses on those issues with design implications. Issues to be discussed include, but are not limited to:

- Sleep
- Clothing
- Exercise
- Medical Support
- Personal Hygiene
- Food Preparation
- Group Interaction
- Habitat Aesthetics
- Outside Communications
- Recreational Opportunities
- Privacy and Personal Space
- Waste Disposal and Management
- Scheduling and Workload Assessment
- Onboard Training and Task Preparation

Objectives:

The objectives of our discussion are to develop a list of critical research issues and to explore possible avenues to resolve problems requiring design solutions.

Panel Members:

Captain Sidney M. Blair, MC, USN
Executive Officer
Naval Hospital Oakland
(a Naval psychiatrist experienced in psychological adaptation to extreme environments)

Gerald P. Carr, (Col. USMC ret.)
Vice-President
CAMUS, Inc.
(the Commander of Skylab 3--the longest-duration U.S. manned space mission, currently a consultant to aerospace firms concerning space station design)

William K. Douglas, MD, (Col. USAF ret.)
Special Fellow
McDonnell Douglas Astronautics Co.
(the original flight surgeon for Project Mercury, currently engaged in space station design studies)

Captain Vance Gilstrap, USAF
Space Division
(a human factors specialist working in the Manned Spaceflight Engineer Program)

William E. Haynes
Manager
Manned Spaceflight Support
The Aerospace Corp.
(a fighter pilot and astronaut candidate early in his career, currently provides technical support and consultation to the MSE Program)

Captain Brian Shoemaker, USN
U.S. Space Command
(an aviator, Antarctic winter-over leader, and former Commander of the Naval Support Force Antarctica; currently representing the Navy at the Unified Space Command, Colorado Springs)

Captian John L. Sullivan, USN
U.S. Space Command
(a former submarine commander currently representing the Navy at the Unified Space Command)

James A. Winters
MITRE Corp.
(a psychologist conducting classified research concerning human responses to stressful working environments)

Jack Stuster, Ph.D. (Panel Chairman)
Senior Scientist
Anacapa Sciences, Inc.
(a behavioral scientist with operations-level experience in extreme environments and a scientific interest in design issues associated with isolation and confinement; Project Director of a recent NASA space station habitability study)

An Animal Model of Sustained Operations and Continuous Performance

G. Jean Kant and John R. Leu

Department of Medical Neurosciences, Walter Reed Army Institute of Research

Abstract

An animal model is described that should prove useful for investigating the effects of chronic stress on physiology and behavior and for testing the effects of potential performance enhancers.

Rats were housed around the clock in operant boxes; food was earned by bar pressing and intermittent footshock could be avoided or escaped by pulling a chain. Control animals were housed similarly but no footshock trials were presented. In the stressed group, rats were presented with a shock trial at approximately 5 min intervals around the clock for up to 2 weeks. All food lever presses and chain pulls were recorded by a PDP8 computer which collected responses in 1 hr bins.

Control rats demonstrated a marked circadian rhythm in lever pressing for food, while stressed rats decreased overall responding for food and also showed disrupted circadian patterns of eating. Water intake was similar in control and stressed groups.

Adrenal weights were increased and thymus weights decreased in stressed as compared to control groups. These changes in organ weights are typical of chronic stress states. Plasma corticosterone, a major rat stress hormone, was elevated in stressed rats after 1 to 7 days in the paradigm, but levels returned to control levels by day 14.

Introduction

Stress is harmful to physical and mental well-being. Chronic stress is thought to be a significant contributing factor to the development of heart disease, gastrointestinal disorders (e.g. ulcers), in decreased resistance to infection via suppression of immune function and in the genesis of depression and other mental illness (1,7,8). Soldiers must perform in an environment that is inherently stressful. In addition to fear, other factors such as sleep disruption, fatigue, extremes of temperature and chemical agents that can interact with and exacerbate the effects of psychological stress are likely to also be present on the battlefield. In past wars, battlefield stress has caused both acute and delayed psychiatric stress causalties. The Vietnam war, for example, produced a number of highly publicized "post-traumatic stress syndrome" casualties. It is likely that future battlefields will produce more intense confrontations with ever increasing firepower and individual isolation. The individual combat soldier and his ability to function in a sustained stressful environment is a key element in the success of the modern army. Clearly, improved stress management via both non-pharmacological and pharmacological means would help conserve the fighting strength acutely and improve the long-term mental health outlook for returning veterans.

However, in order to design new stress management strategies, particulary those involving the use of drugs (e.g. performance enhancers, anti-anxiety drugs, sleep aids etc.), much more information is required regarding the effects of acute and chronic stress on brain biochemistry as well as on physiology and behavior. The work of our laboratory is directed at providing some of this needed information. For that purpose, we have developed and characterized various models of acute and chronic stress (2-5).

The model described in this report incorporates sleep disruption, chronic stress and measures of performance and should prove useful for these studies.

Methods

Experimental Procedures and Apparatus

On day 1, rats were placed in standard operant cages set inside acoustical boxes. A fluorescent house light was on between 0600 and 1800 hrs. Cages were equipped with a triple cue light, a sonalert, lever for food pellet delivery and a ceiling chain. The floor consisted of parallel metal bars connected to a scrambled programmable shocker wired to deliver 0.16, 0.32, 0.65, 1.3 or 2.6 mA of electrical current. All equipment was controlled by and all responses were recorded by a PDP8 computer programmed in SKED. Responses were collected in 1 hr bins.

For the first 3 days, no shocks were delivered to any rats. Rats learned to bar press for food pellets. On day 4, some of the rats (stressed) were trained to pull the ceiling chain to escape footshock delivered and controlled by the experimenter via a hand switch. Once the chain pull response was learned, the hand switch was disconnected and shock presentation was automatically controlled by the PDP8. A shock trial commenced with 5 sec of a warning light followed by a 5 sec tone. Then 5 sec each of escalating shock intensities (0.16, 0.32, 0.65, 1.3 and 2.6mA) followed. The shock sequence was terminated at any point following a ceiling chain pull or shock was terminated after 5 sec at 2.6mA. Until 35 avoidances or escapes were recorded, shock trials were presented at approximately 1 min intervals. Following 35 responses, the intertrial interval was set to 5 min on a variable interval schedule. Mechanical relays prevented any possibility of power failure induced sticking of the shocker in the "on" position.

On day 5, following the first 24 hr of shock trials, rats exceeding a 75% escape from escapable trials criterion were placed on a modified schedule in which 10% of the trials were inescapable, i.e. chain pull was without effect. If the performance of an individual rat degraded on the modified schedule such that less than 75% of escapable trials were escaped, the rat was placed back on the original schedule in which all trials were escapable. Control rats continued to live in cages without any shock trial presentations. Different experiments were performed in which rats were sacrificed at different times following shock trial initiation, i.e. after 1,2,3,4,7 and 14 days of shock trials.

Tissue preparation and Assays

Rats were sacrificed by decapitation. Trunk blood was collected, processed and the plasma stored frozen until assayed for corticosterone by radioimmunoassay (6). The adrenal glands, spleen and thymus were dissected free and weighed.

Results

As shown in Fig 1, daily bar pressing for food increased from day 1 to day 3. After day 4, responding for food remained relatively constant in controls. Shock trial onset on day 4 reduced responding for food in the stress groups. Bar pressing for food gradually returned to control rates . Control rats displayed marked circadian rhythms in bar pressing for food, but this rhythm was degraded in stressed rats (Table 1). Stressed rats had a higher percentage of responses during the light hours as compared to control rats. This change in percentage was primarily due to decreased dark responding in stressed rats combined with a slight increase in light responding compared to controls.

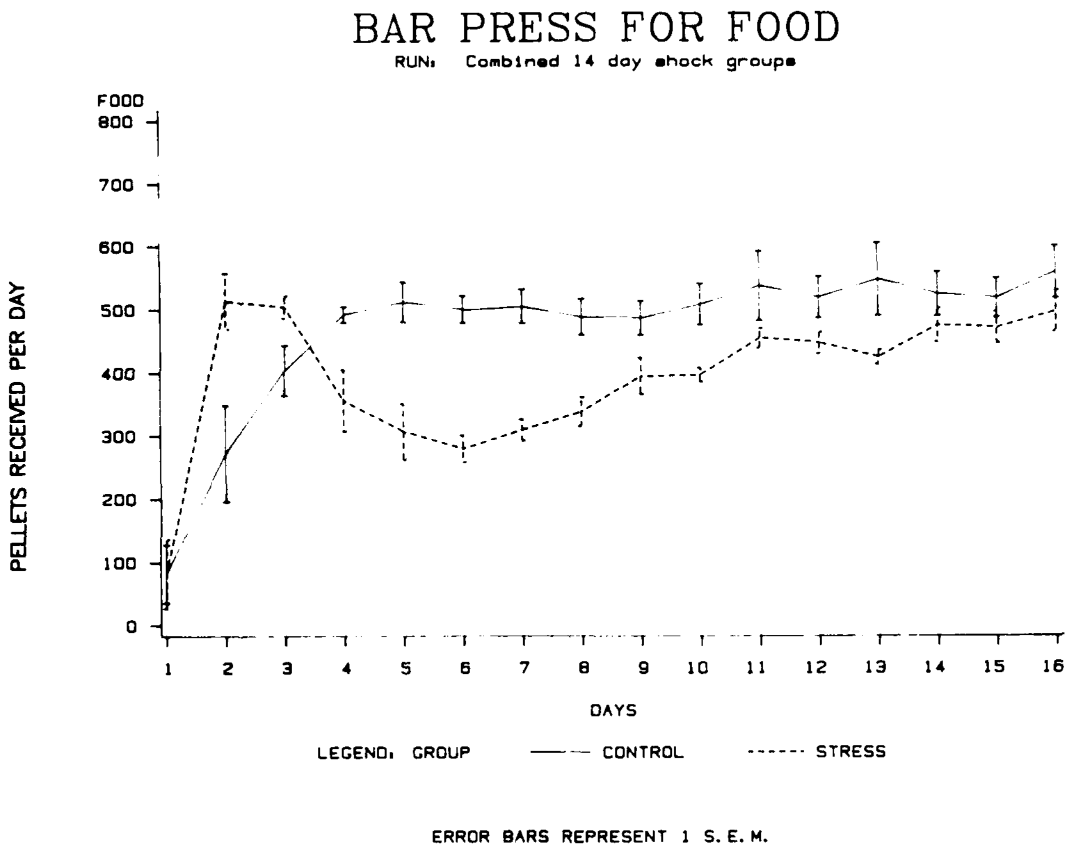


Fig 1. Responding for Food

TABLE 1. Bar Presses for Food

	Light Responses	Dark Responses	% Light
Controls	76 ± 6	434 ± 18	15.4 ± 1.2
Stressed	92 ± 6	311 ± 21	23.1 ± 1.2

The average weight gain for two 14 day shock trial experiments was: 111 grams for controls and 70 grams for stressed rats. Water intake for controls and stressed rats was similar.

The average shock level at which rats avoided or escaped changed over the course of days and also varied with time of day. With avoidances arbitrarily valued at 0, shock escapes at levels 1 through 5 valued at 1 to 5 and no escapes valued at 6, the mean daily shock level decreased from day 5 to day 16 from 4.3 to 3.3. In addition, the lowest mean shock level generally was seen at the midpoint of the dark cycle (midnight) and the highest mean shock escape level was seen during the midportion of the light cycle, i.e. rats escaped earlier in the shock sequence during the dark hours.

As shown in Table 2, adrenal weights were increased and thymus weights decreased

following 14 days of shock trials. Plasma corticosterone levels for various durations of shock trials are shown in Table 3. Initially, stressed rats had high corticosterone levels, but values returned to control levels after 14 days of stress.

TABLE 2. Organ Weights after 14 Days of Shock Trials
Wet weight (mg)

	Controls	Stressed	% Change
Adrenal glands	39.4 ± 3.6	46.8 ± 3.7	+19
Thymus	858 ± 72	578 ± 45	-33
Spleen	888 ± 63	806 ± 52	-9

TABLE 3. Plasma Corticosterone (ug/100ml)

Shock trial Days	Controls	Stressed
1	1.9 ± 0.8	8.7 ± 4.7
2	1.0 ± 0.2	13.4 ± 4.3
3	1.0 ± 0.1	6.9 ± 2.5
4	1.6 ± 0.1	6.2 ± 3.8
7	1.3 ± 0.2	2.7 ± 0.8
14	1.6 ± 0.4	1.2 ± 0.2

Discussion

The model described in this report appears to be suitable for studies of the effects of sustained stress on brain biochemistry, neuroendocrinology, physiology and performance. The paradigm is generally stressful as shown by initially increased levels of plasma corticosterone, food intake reduction, adrenal hypertrophy, decreased thymus weight and disruption of circadian rhythms. On the other hand, all rats tolerated the stressful environment for the two weeks of the study, continuing to eat, gain weight and groom. Once the procedures were learned, most rats escaped greater than 75% of escapable trials and continued this performance for extended periods of time. Differences in escape performance among individual rats, over the course of 14 days and over the course of 24 hrs suggest that this measure of performance may be a useful index of performance that can be assessed following various interventions, e.g. following no shock naps inserted into the schedule or following administration of various drugs.

Acknowledgements

The authors appreciate the excellent technical assistance provided by Jack Delaney, Steve Shuie, Joe Campbell, Hugh Jarrard, David Jarrard, SSG Edward Ramsey, SGT Leigh Landman-Roberts, SGT Terry Eggleston, Clyde C. Kenion, Golden C. Driver, Peter A. Kant and Aaron P. Kant.

Disclaimers

1. The views of the authors do not purport to reflect the position of the Department of the Army or the Department of Defense (para 4-3, AR 360-5).

2. Research was conducted in compliance with the Animal Welfare Act, and other Federal statutes and regulations relating to animals and experiments involving animals and adheres to principles stated in the Guide for the Care and Use of Laboratory Animals, NIH publication 85-23. All procedures were reviewed and approved by the WRAIR Animal Use Review Committee.

References

1. Goldberger, L. and Bresnitz, S. (Ed.). (1982). Handbook of Stress: Theoretical and Clinical Aspects New York, The Free Press.
2. Kant, G.J., Bunnell, B.N., Mougey, E.H., Pennington, L.L., and Meyerhoff, J.L. (1983) Effects of repeated stress on pituitary cyclic AMP and plasma prolactin, corticosterone and growth hormone in male rats. Pharmacology, Biochemistry & Behavior 18 967-971.
3. Kant, G.J., Lenox, R.H., Bunnell, B.N., Mougey, E.H., Pennington, L.L. and Meyerhoff, J.L. (1983). Comparison of stress response in male and female rats: pituitary cyclic AMP and plasma prolactin, growth hormone and corticosterone. Psychoneuroendocrinology 8 421-428.
4. Kant, G.J., Mougey, E.H., Pennington, L.L. and Meyerhoff, J.L. (1984). Graded footshock stress elevates pituitary cyclic AMP and plasma Beta-endorphin, Beta-LPH, corticosterone and prolactin. Life Sciences 33 2657-2663.
5. Kant, G.J., Eggleston, T., Landman-Roberts, L., Kenion, C.C., Driver, G.C., and Meyerhoff, J.L. (1985). Habituation to repeated stress is stressor specific. Pharmacology, Biochemistry & Behavior 22 631-634.
6. Mougey, E.H. (1978) A radioimmunoassay for tetrahydrocortisol. Anal. Biochem. 91 566-582.
7. Sachar, E.J. (1980) Hormonal changes in stress and mental illness. In D.T. Krieger and J.C. Hughes (Ed.), Neuroendocrinology (pp177-184). Sunderland, Mass, Sinauer Associates.
8. Selye, H. (1976). Stress in Health and Disease Boston, Butterworths.

Animal Model for Predicting the Relationship Between Behavioral Incapacitation and Exposure to Organophosphorus Compounds

James A. Romano, Jr. and Mary Z. Mays
US Army Medical Research Institute of Chemical Defense

Abstract

Data from rats and guinea pigs on tests of rotorod, locomotor activity, tail flick, grip strength, shuttle avoidance, and one-way avoidance performance following exposure to GA, GB, GD, or VX were used to model the likelihood of behavioral incapacitation resulting from exposure to these organophosphorus compounds. The proportion of animals behaviorally incapacitated was analyzed as a function of dose and plotted relative to the lethality function. These data demonstrated the extreme toxicity of organophosphorus compounds and the likelihood that even sublethal doses of these compounds will cause significant behavioral incapacitation in those exposed. It is proposed that this approach to describing the relationship between behavioral incapacitation and exposure level could be used to model the probability of combat ineffectiveness induced by exposure to organophosphorus compounds.

In preparing for battle, it is important for commanders to be able to predict the number of casualties created by a particular weapon or type of combat. It is equally important in predicting the combat effectiveness of a unit, to be able to predict the number of soldiers who will be behaviorally incapacitated by a particular weapon or type of combat. A model is needed which will describe the relationship between sublethal exposure to organophosphorus compounds, such as GA, GB, GD, and VX and behavioral incapacitation. Behavioral toxicity data from independent investigations conducted in two laboratories, with two species of rodent, four organophosphorus compounds, and six behavioral tests, were used to develop an animal model of that relationship.

While assigned to the USAF School of Aerospace Medicine, CPT(P) Romano tested rats on rotorod, locomotor activity, tail flick, grip strength and shuttle avoidance tests, 30-40 minutes after subcutaneous injection of GD. Behavioral incapacitation on these tests was defined as performance that was $\pm$ 1.96 standard deviations from the control group mean. Working independently at the USA Medical Research Institute of Chemical Defense, CPT(P) Mays tested rats and guinea pigs on a one-way avoidance test 25 minutes after subcutaneous injection of GA, GB, GD, or VX. Behavioral incapacitation on this test was defined as

performance that was 60% or less of baseline performance. Together these tests comprise a battery of tests of behaviors ranging from simple reflexive ones to complex cognitive ones.

In Figures 1-3, the probability that an animal will meet the behavioral incapacitation criterion is shown as a function of dose (expressed as a proportion of the LD50 dose) and plotted relative to the lethality function. Figure 1 shows the results of analysis of the performance of rats on five behavioral tests following exposure to GD. Note that the ED50 on all behavioral tests was predicted to be less than the LD10.

Figure 1

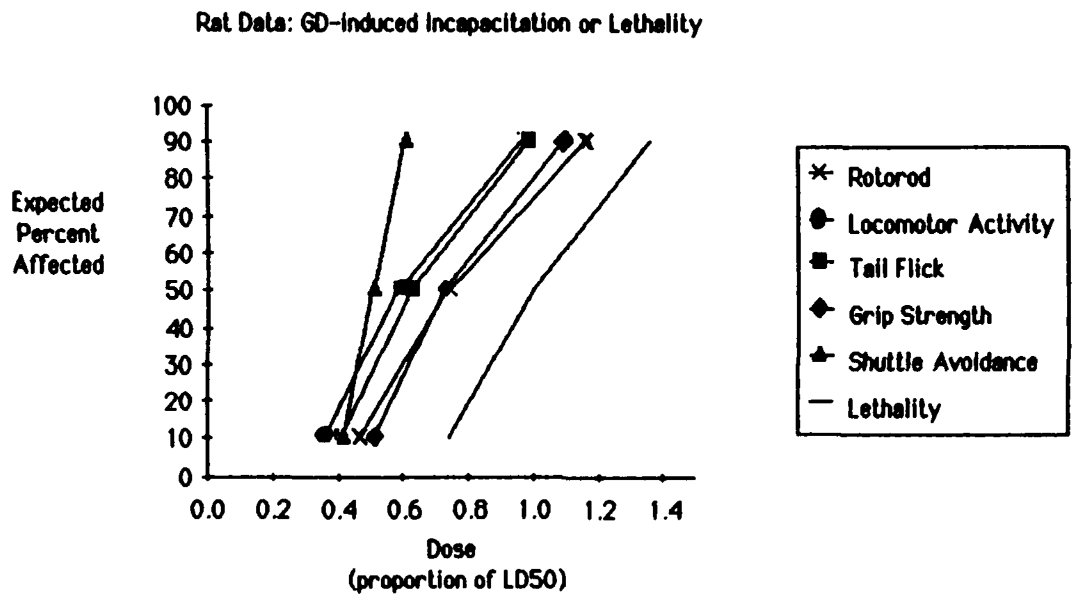


Figure 2 shows the results of analysis of the performance of guinea pigs on the one-way avoidance test following exposure to GA, GB, GD, or VX. In these experiments, the ED50 on the behavioral test is predicted to be less than the respective LD1 for each agent.

Figure 2

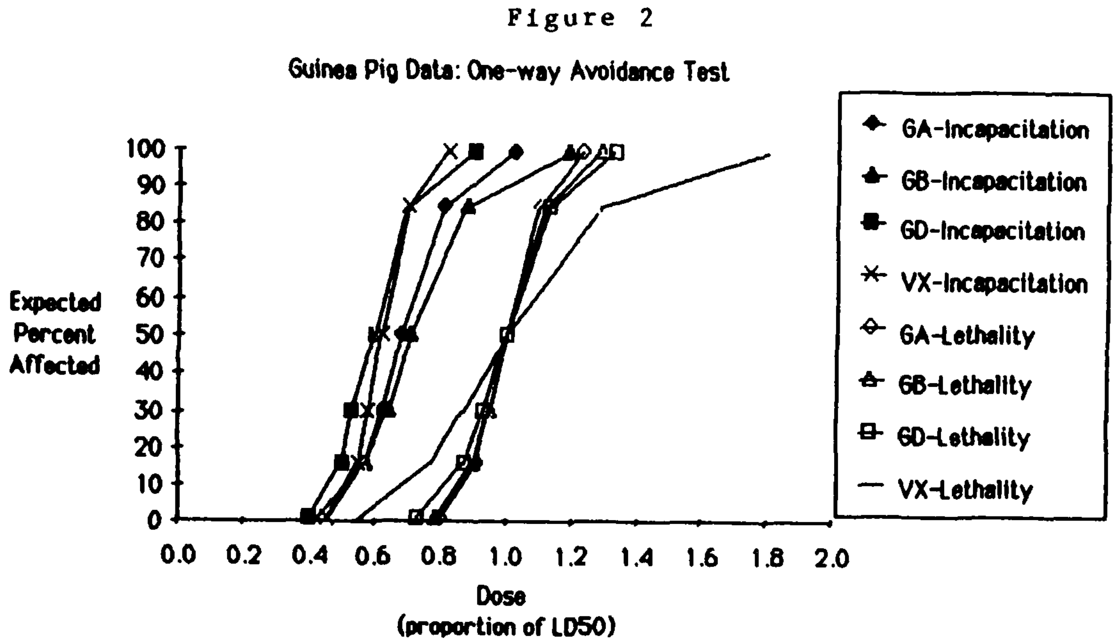
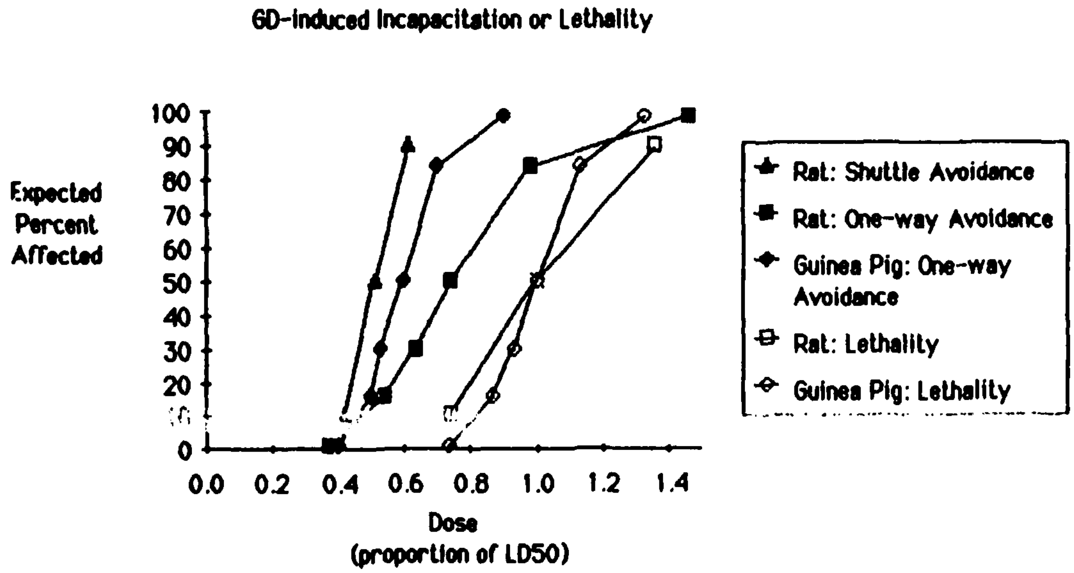


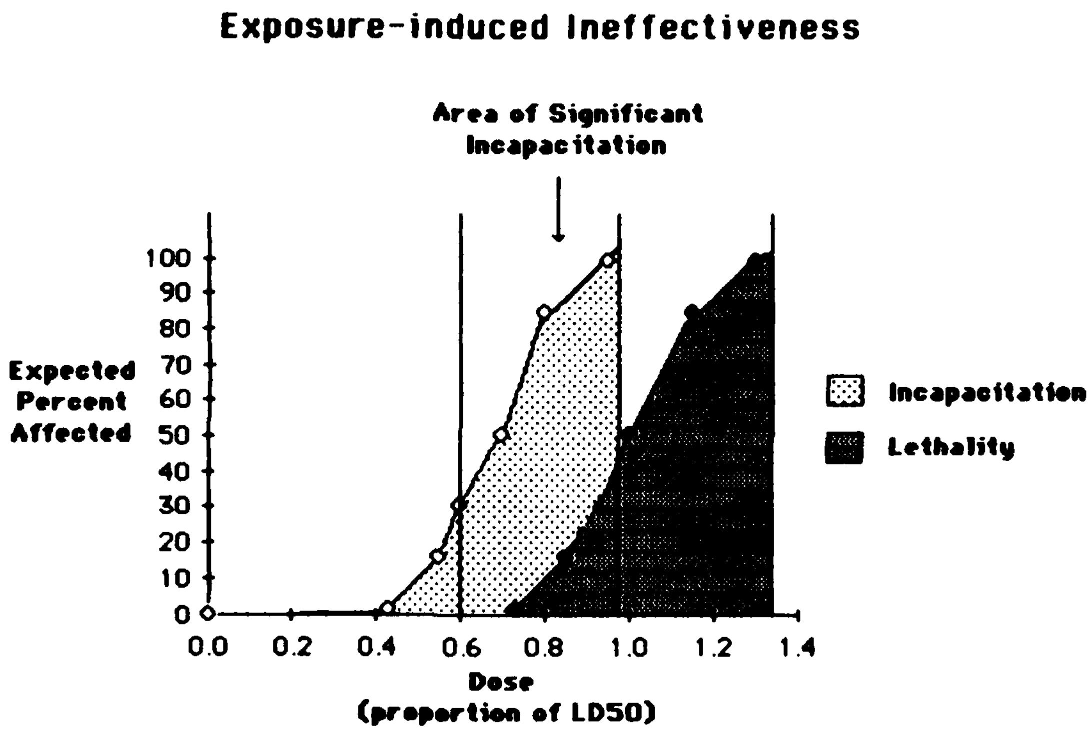
Figure 3 presents a comparison of the performance of rats and guinea pigs on two types of avoidance tests followng exposure to GD.

Figure 3



These cumulative probability ogives demonstrated the extreme toxicity of organophosphorus compounds and the likelihood that even sublethal doses of these compounds will cause significant behavioral incapacitation in those exposed. Given the large body of behavioral toxicology data available on organophosphorus compounds, this approach could be used to predict the theoretical relationship between behavioral incapacitation and exposure level as shown in Figure 4.

Figure 4



There are four advantages to this model. It employs the commonly used method of probit analysis. Data from a wide variety of behavioral tests can be converted easily for inclusion in the model, providing an extremely broad database. It predicts the likelihood of any given level of behavioral incapacitation across a range of doses. It describes the likelihood of behavioral incapacitation relative to lethality. Ultimately, this type of model may prove useful in predicting the probability of exposure-induced combat ineffectiveness and in assessing the ability of pretreatment and antidote drugs to attenuate or eliminate performance deficits.

Disclaimers. The experiments reported here were conducted according to the Guide for Care and Use of Laboratory Animals (1978) as prepared by the Committee on Care and Use of Laboratory Animals, National Research Council, DHEW Publication No. (NIH) 80-23. The opinions or assertions contained herein are the private views of the authors and are not to be construed as reflecting the views of the Department of the Army or the Department of Defense.

Effects of Depletion of Cerebral Somatostatin on Passive Avoidance Conditioning in Rats

M. D. Matthews,[1] G. R. Sessions,[1] C. Bakit,[2] N. Swerdlow,[2] A. K. Mattthews[1] and G. F. Koob[2]

[1]United States Air Force Academy

[2]Division of Preclinical Neurosciences and Endocrinology
Scripps Clinic and Research Foundation

Abstract

The current study examined the effects of depletion of cerebral somatostatin on passive avoidance conditioning in rats. Ten male Wister rats were depleted of somatostatin via intracerebroventricular (i.c.v.) injections of cysteamine, and their performance on a step-through passive avoidance task was compared to that of rats receiving i.c.v. saline injections. Results revealed no significant effects of somatostatin depletion on passive avoidance conditioning. Viewed in conjunction with findings from other studies, the role of somatostatin in mediating aversively motivated behaviors is questioned.

Recently, the role of various neurochemicals, particularly acetylcholine (ACh), have been investigated as possible factors in the development of senile dementia of the Alzheimer's type (Ferrier, et al., 1983). While ACh has been the most actively researched substance, a neuropeptitide, somatostatin, has also been implicated in a number of dementias, including Alzheimer's disease (Roberts, Crow, & Polak, 1985), Parkinson's disease (Epelbaum, et al., 1983), and Huntington's disease (Aronin et al., 1983). Somatostatin's role in Alzheimer's disease is particularly intriguing. Roberts et al (1985) report that at least some somatostatin neurons are specifically affected by the Alzheimer's disease process, and they further conclude that somatostatin depletion is the primary causative factor in the disease with ACh depletion playing a secondary role.

While the relationship between brain somatostatin levels and dementia has been based to date in large part on post-mortum comparisons between brains of dementia patients and non-diseased controls, some preliminary investigations utilizing animal subjects have also investigated the role of somatostatin in the acquisition and retention of a variety of behaviors. Intracerebral injections of somatostatin have been shown to inhibit the extinction of active avoidance responding (Vecsie, Bollock, & Telegday, 1983a), increase ambulatory behavior (Vecsie, Bollock, & Telegday, 1983b) and activity levels (Bollock, Vecsei, & Telegday, 1983; Vescie, et al., 1984), and block ECS-induced amnesnia (Vecsie, Bollock, & Telegday, 1983c) in rats. Small increases in brain somatostatin levels have been linked to increases in intracranial self-stimulation, but larger increases apparently decrease such behavior (Vecsie, Schwarzburg, & Telegday, 1982). Increased somatostatin levels do not affect

acquisition of a T-maze response (Vecsie et al., 1984). In contrast, depletions of brain somatostatin levels have been shown to either reduce activity (Vecsie et al., 1984) or to have no effect on activity (Sessions, Matthews, Bakit, Swerdlow, & Koob, 1985) in rats, and to have no impact on acquisition of active avoidance or spontaneous alternations in a Y maze (Sessions et al., 1985). Finally, Bakit and Swerdlow (1986) report that depletion of somatostatin results in a passive avoidance decrement.

The current study is part of a series of experiments designed to further examine the role of somatostatin depletion on a variety of tasks assessing learning and memory. Specifically, the role of somatostatin in mediating passive avoidance performance was assessed by depleting rats of somatostatin prior to passive avoidance training. Somatostatin was depleted by multiple injections of cysteamine HCl, which has been shown to reduce somatostatin-14 (SS-14) immunoreactivity in cortex, hippocampus, and and hypothalamus for up to seven days (Bakit & Swerdow, 1986; Bakit, Benoit, & Bloom, 1983) following repeated injections. Twenty-four hour retention of a step through passive avoidance task was tested in rats following a regimin of three daily i.c.v. administrations of cysteamine HCl.

Method

Subjects. For the passive avoidance study, twenty male Wister rats were used. They averaged 285g in weight at the beginning of the study, and had grown to an average weight of 330g at time of sacrifice. They were housed individually with free access to standard laboratory rat chow and water. Lighting was maintained on a 12-12 light-dark cycle with light onset at 0800 h.

Additionally, 11 male Wister rats were used in a study designed to measure somatostatin depletion levels two days following their last cysteamine HCl injection. These animals were the same age as the passive avoidance animals, and were treated similarly.

Apparatus A two-compartment shuttlebox consisting of two 20.3 x 10.2 x 20.3 cm chambers, one white and one black, connected by a white 8.9 x 6.3 x 7.6 cm passageway, was used. The floor of the apparatus was constructed of steel rods 3.8 cm apart, and two Plexiglas guillotine doors separated the two compartments from the passageway. The hinged lids of the two chambers consisted of Plexiglas. Digital logic manufactured by BRS Foringer was used to program stimulus events, and a BRS/LVE Model SGS-003 shocker scrambler was used to electrify the grid floor of the apparatus.

Procedure. Under sodium pentobarbital anesthesia, cannulla guides were permanently implanted into the lateral ventricles (Kobb, et al., 1984) of all subjects (position counterbalanced across left and right lateral ventricles).Two weeks were allowed for recovery from the surgery to cysteamine injections and behavioral testing. During this period the animals were handled

for approximately three minutes a day for seven days prior to drug injection.

Cysteamine HCl (350 ug in 2 ul phosphate buffered saline) was injected into the lateral ventricles once each day for three consecutive days. Control animals received injections of the vehicle only. Animals given passive avoidance training were given two days to recover before behavioral testing was initiated. The 11 animals used to investigate the degree of somatostatin depletion two days following the injection regimin were sacrificed 48 hours following their last of three daily injections.

Behavioral testing consisted of placing each subject in the white compartment of the apparatus with the guillotine doors open thus allowing it to explore the entire apparatus for a single 10 minute period. Approximately two hours later, the subject was again placed into the white compartment of the apparatus, this time with a guillotine door in place to prevent it from crossing into the black compartment. Five seconds following placement in the white compartment, the door was raised and the rat was free to move to the black compartment. Immediately upon entrance into the black compartment, the guillotine door separating the black compartment from the passageway was lowered, and 3.0 seconds of 1.0 mA scrambled shock was delivered to the subject. The subject was left in the black compartment for 30 seconds following the shock, then it was removed and returned to its home cage. Twenty-four hours later, the subject was replaced into the white compartment, the guillotine door raised five seconds later, and latency to cross into the black compartment was recorded.

Following either the passive avoidance trials (passive avoidance group) or 48 hours following their last injection (marker group), the subjects were sacrificed by decapitation, their brains quickly removed, and the hypothalamus, hippocampus, and cortex dissected on ice. The dissected portions of the brains were immediately frozen in dry ice to await neurochemical analysis. Somatostatin-like immunoreactivity for SS-14 was measured in the whole cortex, hypothalamus, and hippocampus by radioimmunoassay using specific antisera (Bakit et al, 1983).

Results and Discussion

Table 1 shows mean regional values of SS-14 (expressed in pmols/100 mg protein) at 48 and 96 hours following the last injection, for cortex, hippocampus, and hypothalamus. For the 48 hours data, the cortex and hypothalamus showed significantly lower levels of somatostatin (t=2.8 and 9.4; df= 9; $P < .05$, respectively) for animals injected with cysteamine compared to values of animals given saline injections. Ninety-six hours following the last injection, only the hypothalamus remained significantly depleted of somatostatin relative to controls (t=7.14; df= 16; $p < .01$).

The mean cross-over latencies for the punishment trial of passive avoidance training were 5.6 seconds and 17.6 seconds for the cysteamine injected and saline injected animals, respectively. However, analysis by Mann Whitney U test indicated no statistically signficant difference between the two groups

(U=57; p > .05). The medians for the cysteamine and saline injected groups were 4.0 and 5.5, respectively. On the passive avoidance trial, the cysteamine and saline injected groups had mean cross-over latencies of 198.81 and 175.36 seconds, respectively. This difference was not statistically significant (Mann-Whitney U = 58; p > .05).

The results of the current study fail to replicate prelimimary findings of Bakit and Swerdlow (1986), who did obtain a passive avoidance decrement for animals depleted of somatostatin. There is at least one difference in the procedures between the current study and that of Bakit and Swerdlow that might account for the discrepant findings. Whereas the current study used a series of three cysteamine injections, Bakit and Swerdow used a regimin of four injections. The three injection procedure was used in the current study because significant mortality was observed previously when a four injection regimin was used. An additional injection may be necessary to result in a passive avoidance decrement. Nevertheless, it is interesting to note that the percentages of somatostatin depletion reported were similar in both experiments.

Viewed in conjunction with earlier findings that somatostatin depletion has no effect on active avoidance acquisition (Sessions et al, 1985), the results from the current study suggest that somatostain depletion, at least at depletion levels obtained in these studies, has little effect on acquisition or retention of aversively motivated tasks. The role of somatostatin depletion on more complex behaviors, for example discrimination learning or ongoing operant choice behavior, is not yet known.

References

Aronin, N., Cooper, P. E., Lorenz, L. J., Bird, E. D., Sager, S. M., Leeman, S. E., & Martin, J. B. (1983). Somatostatin is increased in the basal ganglia in Huntington's disease. Annals of Neurology, 13, 519-526.

Bakit, C., & Swerdlow, N. (1986). Behavioral changes following central injections of cysteamine in rats. Brain Research, in press.

Bakit, C., Benoit, R., & Bloom, F. E. (1983). Effects of cysteamine on pro-somatostatin related peptides. Regulatory Peptides, 6, 169-177.

Bollok, I., Vescei, L., & Telegdy, G. (1983). The effects of interaction between propranolol and somatostatin on the active avoidance behavior, open-field activity and electroconvulsive shock-induced amnesnia in rats. Neuropeptides, 3, 263-270.

Epelbaum, J., Ruberg, M., Moyse, E., Javoy-Agid, F., DuBois, B., & Agid, Y. (1983). Somatostatin and dementia in Parkinson's disease. Brain Research, 278, 376-379.

Ferrier, I. N., Cross, A. J., Johnson, J. A., Roberts, G. W., Crow, T. J., Corsellis, J. A. N., Lee, Y. C., O'Shaughnessy, D., Adrian, T. E., McGregor, G. P., Baracese-Hamilton, A. J., & Bloom, S. R. (1983). Neuropeptides in Alzheimer's type dementia. Journal of Neurological Sciences, 62, 159-170.

Kobb, G. H., Swerdlow, N., Seeligson, M., Eaves, M., Sutton, R.,

Rivier, J., & Vail, W. (1984). Effects of alpha-flypenthioxol and naloxone on CRF-induced locomotor activation. Neuroendocrinology, 39, 459-464.

Roberts, G. W., Crow, T. J., & Polak, J. M. (1985). Location of neuronal tangles in somatostatin neurons in Alzheimer's disease. Nature, 314, 92-94.

Sessions, G. R., Matthews, M. D., Swerdlow, N., Bakit, C., & Koob, G. F. (1985). A comparison of cerebral acetylcholine or somatostatin depletion on avoidance conditioning, activity, and working memory in rats. Neuroscience Abstracts, 11, 1113.

Vescei, L., Bollock, I., & Telegdy, G. (1983a). The effect of linear somatostatin on the active avoidance behaviour and open-field activity on haloperidol, phenoxybenzamine and atropine pretreated rats. Acta Physiologica Hungarica, 62, 205-211.

Vescie, L., Bollock, I., & Telegdy, G. (1983b). Comparative studies with cyclic and linear somatostatin on active avoidance behaviour and open-field activity in rats. Acta Physiologica Hungaria, 61, 43-49.

Vescie, L., Bollock, I., & Telegdy, G. (1983c). Intracerebroventricular somatostatin attenuates electroconvulsive shock-induced amnesia in rats. Peptides, 4, 293-295.

Vescie, L., Schwarzberg, H., & Telegdy, G. (1982). The effect of somatostatin on the self-stimulation of rats. Neuroendocrinological Letters, 4, 37-41.

Vescie, L., Kiraly, C., Bollock, I., Nagy, A., Varga, J., Penke, B., & Telegdy, G. (1984). Comparative studies with somatostatin and cysteamine in different behavioral tests on rats. Pharmacology Biochemistry & Behavior, 21, 833-837.

Table 1

Mean cerebral somatostatin-14 like immunoreactivty levels expressed in pmols/100 mg protein, for cysteamine injected (CYS) and saline injected (SAL) animals.

		Cortex	Hippocampus	Hypothalamus
48 hrs	CYS	16.0	11.0	74.2
	SAL	26.0	17.8	186.4
96 hrs	CYS	35.3	20.6	64.5
	SAL	27.1	18.5	131.8

Spatial memory in non-human primates[1]

Timothy F. Elsmore and John K. Parkinson
Walter Reed Army Institute of Research

Roger L. Mellgren
University of Oklahoma

Abstract

A procedure for assessing spatial memory in non-human primates is described. A 3 X 3 matrix of red squares was displayed on a video monitor, with one square marked by a green circle. In the first experiment, this sample display remained on for 5 sec, followed by a blank screen for a period ranging from 0 to 32 sec. Following the termination of this retention interval, nine red squares were displayed, and the animal received food for touching the square that previously had been marked with a circle. Smooth accuracy/delay functions were observed with the animals performing at about 50% correct at 32 sec delays. Requiring the animals to respond to the corners of the screen to initiate the delay reduced accuracy. Additional applications of this technology are discussed.

Introduction

The validity of animal models of human performance is usually assessed by evaluating the degree to which they predict human performance on similar tasks. The key term in this statement is **similar.** In my previous paper at this conference (Elsmore 1984), I discussed structural and functional approaches to animal modeling of human behavior, emphasizing the value of functional models in which particular behavioral functions (e.g. memory, learning, psychomotor coordination) are assessed.

Structural approaches to behavioral modeling, on the other hand, emphasize formal similarity between animal and human testing procedures. In other words, the animal test **looks like** a task humans might be required to perform in some practical situation. While such tests frequently have great face validity, interpretation of results may be difficult, since there may be little functional validity. It is possible, however to strike a happy medium. Animal testing situations may have both functional and structural similarities to practical human situations.

One difficulty in devising animal tests with good structural properties lies in instrumentation. Typical human testing situations involve highly complex stimulus displays, with many different response alternatives. Simulation of such situations in animal models is frequently constrained by limitations of conventional behavioral testing apparatus. The availability of video touch-screen technology provides the opportunity to free experimenters from the constraints present in typical electro-mechanical and

---

1. In conducting the research described in this report, the investigators adhered to the 'Guide for the Care and Use of Laboratory Animals', as promulgated by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Animal Resources, National Research Council.

The views of the authors do not purport to reflect the position of the Department of the Army or the Department of Defense, (para 4-3, AR 360-5).

projection technologies by providing the opportunity for one-to-one mapping of responses onto a continuous two dimensional stimulus display surface. There have been numerous efforts to approximate this sort of mapping using one-dimensional response surfaces. For example, Cumming and Eckerman (1965) trained pigeons to peck near one extreme of a continuous response strip in the presence of bright illumination, and to peck near the other end in dim illumination. They then determined the position of responses during extinction under conditions of intermediate illumination. Several similar experiments are discussed in a recent review by Bickel and Etzel (1985).

Projection techniques have been used extensively for generation of complex stimulus displays. For example Wright, Santiago, Sands, Kendrick, and Cook (1985) used a slide projection system to study serial list processing by pigeons, monkeys, and people. While these techniques are useful for presentation of complex stimuli, response input is by means of buttons or levers, and thus is inherently limited.

A very similar application to that described in the present paper was done by Blough (1977) who projected symbols on a cathode-ray tube (CRT) mounted on the wall of a pigeon's test cage. A photocell, glued to a pigeon's beak, detected proximity of the beak to one of the symbols being displayed on the screen through synchronization with the sweep circuitry of the CRT, the same principal used by light pens.

Finally, a system nearly identical in function to the present one was reported by a group of researchers from Warner-Lambert/Parke-Davis Pharmaceutical Research (Davis, Tew, and Marriot, 1985). They used a video touch-screen system for evaluating diazepam-induced memory impairments in young and aged monkeys. A matrix of nine white squares was projected on a video monitor, with one alternating white and green. Following a retention interval, the white squares were projected again, and the animals were reinforced for touching the spot that previously flashed green.

The present system was designed for maximum flexibility in terms of generation, modification, and display of images, and sensing of contacts with the screen. A block diagram of the system is shown in Figure 1. The heart of the system is micro PDP11/23 computer that serves as an intelligent controller for the monitor and touch screen. The 11/23 has 512 kbytes of memory and is equipped with a terminal (DEC VT220), a 10Mbyte hard disk, dual 8" floppy disk drives, and a color graphics interface (Parallax Graphics, 640 x 480 pixel resolution). The monitor is a 19" diagonal Mitsubishi model 3910 with a low-persistence phosphor, and a touch sensing system (Carroll Touch Technologies) mounted in the bezel of the monitor. Communication between the touch sensor and the 11/23 is via an RS232 line operating at 9600 baud. The 11/23 is in turn controlled by a micro PDP 11/73 computer (the "host") which uses SKED11 software (Snapper and Inglis, 1985) operating under the RSX11M+ operating system for experimental control and data collection. Communication between the two computers is via two 9600 baud serial lines. Experimental sessions are conducted by attaching a cart containing the video monitor and food pellet dispenser to the front of the animals' living cages. A more complete description is available (Elsmore, Parkinson, and Mellgren, 1986).

Method

Three experimentally niave female Rhesus monkeys served as subjects in these experiments. For several days prior to the initiation of training procedures, Noyes food pellets were substituted for the animals' normal food. Following one to three daily sessions of magazine training in which approximately 100 pellets were delivered, the animals were trained to touch a large (20 cm) square on the screen. An autoshaping

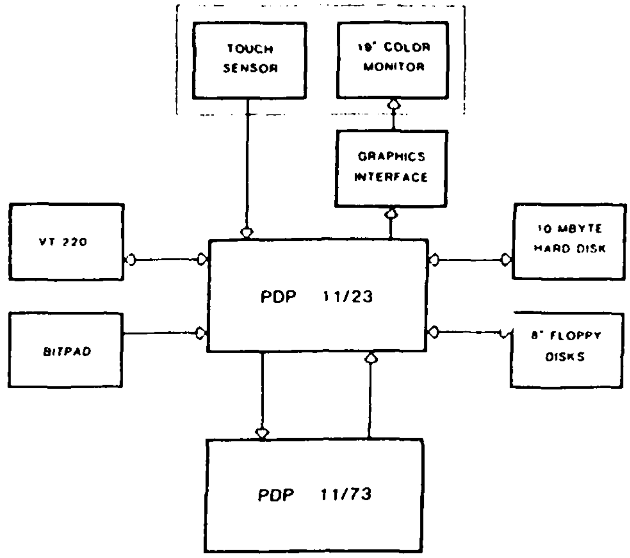


Figure 1. Block diagram of video touch-screen apparatus.

procedure was employed in which the square was displayed for 10 sec, following which the screen was cleared, and a pellet was delivered. If the red square was touched, a pellet was delivered immediately. The intertrial interval was 20 sec. In three of the four animals, this procedure was not effective, and additional manual shaping procedures were required. Once touching the large square was established, the size of the square was reduced in several steps down to 4.5 cm, with approximately 20 pellets at each size.

Once touching the screen was established, more complex discrimination procedures were initiated. Two squares were displayed, one containing a 3.5 cm diameter green circle. The animals were reinforced for touching the square containing the circle. Finally, a spatial memory procedure was initiated in which a sample display containing two squares was presented, with one square one containing the green circle. After 5 sec, the circle was removed, and the animals were reinforced for touching the square which had previously contained the circle. One session was run each day, with the animals earning a maximum of 108 pellets each session. This procedure was designated the zero-delay procedure. When animals achieved 75% accuracy in a daily session, the number of squares was increased up to a maximum of nine. After responding was well-established on the nine-square problem, the delay between termination of the sample display and the onset of the choice display was systematically manipulated in an ascending series, with one session at each of 2, 4, 8, 16, and 32 seconds (single-delay procedure). Following the 32-second delay session, the animals were returned to the zero-delay procedure and baseline performance reestablished. The delay series was then replicated twice.

Two modifications of this procedure were then investigated. In the first modification, all six possible delay values were presented in a random order in a single session, with a maximum of 120 pellets in each session (multiple-delay procedure). Ten to fifteen sessions were run with this procedure. In a second modification, rather than presenting the sample display for a fixed time, the animals were required to touch a square which was randomly presented in one of the four corner of the screen to terminate the sample (corner-touch procedure). Only preliminary data are available for this procedure.

Results

Initial training proceeded in a systematic fashion. Concerns regarding the durability of the glass screen were unfounded. The subjects did not display aggressive behavior towards the screen, the typical response topography being a gentle touch or brushing action with the tips of the fingers. The two-square zero-delay procedure was acquired by all animals within 11 sessions, with progressively fewer sessions required on each problem as the number of alternatives was increased. Two of the three animals required only a single day on each problem when the number of alternatives rose above four.

In both the single- and multiple-delay procedures, systematic functions were found relating accuracy to the sample-comparison delay, with two of the three animals performing considerably above chance (11%) at the longest (32 sec) delay. Performance of all animals was poorly maintained at the longest delay values on the single-delay procedure. The multiple-delay procedure, however, maintained performance well, and produced slightly higher levels of accuracy than the single-delay procedure. Figure 2 shows accuracy for two of the animals pooled across five sessions. The third animal failed to achieve stable performance on 32 second delays in time for this paper.

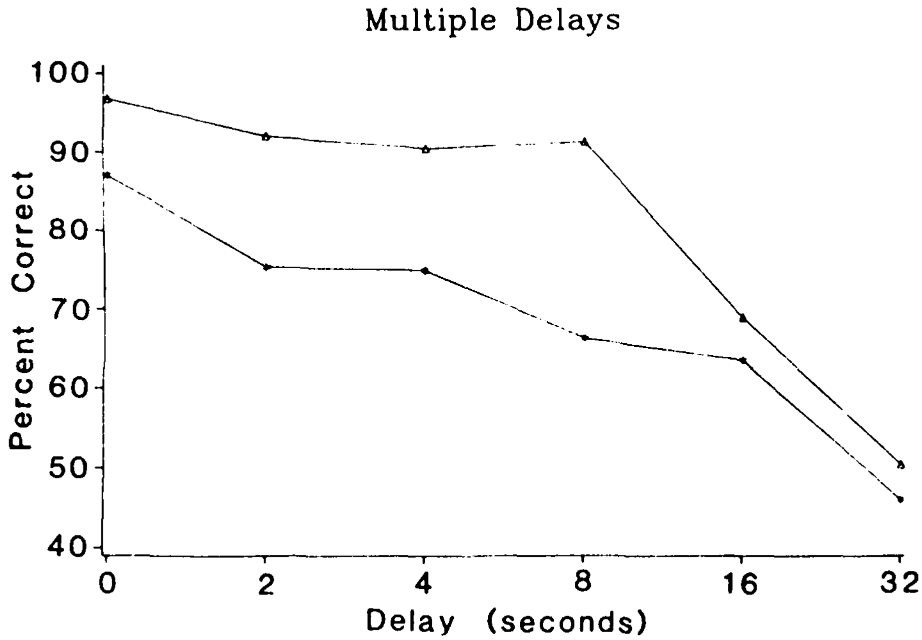


Figure 2. Accuracy of choice in a 9-alternative spatial memory task as a function of the delay between termination of the sample display and onset of the choice display. Each function is for a different monkey.

The rationale for the corner-touch procedure was to investigate the hypothesis that the animals were "bridging the delay" by either holding their hand in spatial proximity to, or visually fixating the correct square. The corner-touch procedure proved to be highly disruptive. In order to establish performance on this procedure, it was necessary to begin with a zero-delay, two-square problem. The two animals run on this procedure required a minimum of one week to achieve 75% accuracy.

Discussion

The purpose of the research presented in this paper was to determine the feasibility of using video touch-screen technology in animal behavior studies, with the goal of providing a balance between structural and functional approaches to animal models of human behavior. The spatial memory paradigm presented here is a single example of a type of experiment that is possible with this new technology, but very difficult with traditional laboratory instrumentation. Spatial memory is a critical component of many human tasks. Until now, there have been few if any controlled laboratory studies of this behavioral function in non-human primates. Animal models that do exist consist largely of mazes. For example, Levy, Kluge, and Elsmore (1983) showed differential actions of centrally and peripherally acting cholinergic drugs on radial-arm maze behavior of mice. The procedures described in this paper will permit an analogous study in primates, both human and non-human.

The infinite variety of images possible using this technology greatly expands the range of experimental questions that may be addressed in the animal laboratory. The ease with which additional images can be generated, and new response areas defined greatly simplifies the implementation of complex experimental procedures. The ability to generate complex stimulus displays should enable experimenters to generate stimulus displays that are more "meaningful" and appropriate for non-human primates than traditional colors, geometric forms, and tilted lines. In cases where species generality of particular research findings is critical, human and non-human primates could even be tested on the same apparatus. The structural identity of animal and human testing situations in such a study would go a long way towards increasing the confidence that the obtained animal data is relevant to the human condition.

References

Bickel, W. K. and Etzel, B. C. (1985). The quantal nature of controlling stimulus-response relations as measured in tests of stimulus generalization. Journal of the Experimental Analysis of Behavior, 44, 245-270.

Blough, D. S. (1977). Photoelectric recording of pigeon-peck responses to computer-driven visual displays. Behavior Research Methods and Instrumentation, 9, 259-262.

Cumming, W. W. and Eckerman, D. A. (1965). Stimulus control of a differentiated operant. Psychonomic Science, 3, 313-314.

Davis, R. E., Tew, T. F., and Marriott, J. G. (1985). CI-933 reverses diazepam-induced memory impairments in young and aged monkeys. Neuroscience Abstracts, 11,187.

Elsmore, T. F. (1984). Animal models of human performance: structural and functional approaches to extrapolating from animal to man. Ninth Psychology in the DOD symposium, Colorado Springs.

Elsmore, T. F., Parkinson, J. K., and Mellgren, R. L. (1986). Video touch-screen stimulus/response surface for use with primates. In preparation.

Levy, A., Kluge, P. B. and Elsmore, T. F. (1983). Radial arm maze performance of mice: acquisition and atropine effects. Behavioral and Neural Biology, 39, 229-240.

Snapper, A. G. and Inglis, G. B. (1985). SKED-11 Software System. State Systems: Kalamazoo, MI.

Wright, A. A., Santiago, H. C., Sands, S. F., Kendrick, D. F., and Cook, R. G. (1985). Memory processing of serial lists by pigeons, monkeys, and people. Science, 229, 287-289.

Efects of Nucleus Basalis Lesions on General Activity in the Rat

G. Rufus Sessions and Charles L. Winstead — US Air Force Academy
Kathryn H. Chantry — University of Colorado

Charles Bakhit and George F. Koob
Scripps Clinic and Research Foundation

Abstract

Home cage activity differentiated into ten discrete behaviors was measured at four time periods of the day in rats with ibotenic acid lesions of the nucleus basalis magnocellularis (NBM) and vehicle controls, using a time sampling observational technique. The lesions of the NBM resulted in a 45 and 44% depletion of choline acetyltransferase in frontal and somatosensory cortex, respectively, while sparing hippocampus. Results of the behavioral testing revealed no consistent enhancement in frequency of occurrence of any of the behaviors associated with hyperactivity nor evidence of overall hyperactivity in either the dark or light. These results indicate that depletion of cortical markers for acetylcholine by lesions of the basal forebrain in rats may be dissociable from the hyperactivity syndrome often observed following lesions of this area.

The ascending cholinergic pathways innervating the cortex from the nucleus basalis magnocellularis (NBM) in the basal forebrain have been implicated in a variety of neuropsychological deficits, including senile dementia of the Alzheimer's type (1,4,13,14,16). An animal model for the study of the functional significance of this neuronal pathway has been used which involves lesioning the cells of the NBM in the ventral pallidum, more specifically, in the area of the substantia innominata (SI) and lateral preoptic area (LPO) of the rat (11). Lesions of the NBM in the rat by means of electrocoagulation or injections of neurochemical toxins have been reported to reliably deplete markers for ACh in cortex (10,11) and to result in disturbances of activity (5), active and passive avoidance (6), and spatial learning and memory (5,8).

Studies of changes in locomotor activity following NBM lesions have revealed a profound and long-lasting enhancement of locomotor activity in tests involving a variety of measurement devices and testing intervals. Measures of locomotor activity in the open field and exploratory head pokes have both been shown to be elevated in lesioned animals (5) . Home cage activity has not been studied, nor has it been determined which components of activity are affected by NBM lesions other than locomotor and exploratory activity in a novel environment.

The purpose of the present experiment was to determine the effects of NBM lesions on general activity of the rat in the home cage environment, using a method that would allow for a qualitative as well as quantitative measurement of activity. Using a time sampling method (2,9) the frequency of occurrence of

ten discrete behaviors was recorded from lesioned and control rats at four different time periods during the light-dark cycle. Results indicated that contrary to expectations, NBM lesions did not result in elevations of overall activity at any of the time periods studied, nor were there consistent patterns of elevations in any of the discrete behaviors measured.

Methods

Subjects

Thirty-two adult male Wistar rats weighing approximately 250 g at time of surgery served as subjects. The animals were housed individually in 25.5 x 19.0 x 15.2 cm plastic cages with wire tops. Corn husk shavings approximately 2.5 cm deep served as bedding. Purina rat chow and water from a ~~standard water from a~~ stainless drinking tube were available ad libitum through the top of the cages. The animals were housed in an air conditioned room under a 12-12 light-dark cycle, with full overhead lighting 0800-2000 h, and very dim lighting 2000-0800.

Surgery

The animals were divided into two groups. Under pentobarbital anesthesia (50 mg/kg, i.p.) sixteen animals received bilateral infusions of ibotenic acid (15 µg/µl, expressed as a salt, infused in a volume of 0.5 µl for 7 min) into the SI and LPO (coordinates from bregma with skull level: AP -0.3, L ± 2.2, DV 8.5 mm ventral to the skull surface). The remaining 16 animals received infusions of the vehicle (0.5 µl phosphate buffer solution, pH 7.4). The subjects were allowed 12-13 days to recover from the effects of surgery before behavioral observations were begun.

Behavioral Testing

Observations of activity were conducted during one-hour test sessions beginning at 2100, 0600, 0900 and 1500 h on three consecutive days. Ten minutes prior to each test session, the observer entered the room housing the animals and seated himself in the center of the room. The animals' cages were arranged in three rows on shelves along two perpendicular walls of the room. The positions of the cages of the lesioned and control animals were arranged symmetrically so that position was counterbalanced across shelves.

During testing, the observer sequentially recorded the behavior of each animal every two minutes until 25 observations were recorded. Observations were made at approximately three sec intervals. Each observation consisted of categorizing the predominant activity of the animal at the moment of observation into one of ten categories. The categories were as follows: Walking, Rearing, Standing, Sniffing (while standing or lying), Grooming, Feeding, Drinking, Digging, Lying With Head Up, and Lying With Head Down. Data collection was computerized and observations were keyed into the keypad of a microcomputer as one or two digit numbers. The primary datum that was analyzed was the frequency of occurrence of each behavior or collections of behaviors at each test session or for combined test sessions conducted in the dark or light portions of the lighting cycle.

Biochemical Methods and Histology

All NBM animals and eight randomly selected control animals were sacrificed by decapitation 38-39 days following surgery,

their brains quickly removed, and hand dissections were conducted to obtain frontal and sensorymotor cortex and hippocampus from each hemisphere. The tissue samples were immediately frozen on dry ice and preserved at -70 C until assayed for levels of choline acetyltransferase (ChAT) in combined hemispheres using a sensitive radiochemical method (3,7).

The remaining brain tissue was fixed in 10% formalin and 40 um thick frozen coronal sections were cut on a sliding microtome and stained with thionine. The results of the ibotenic acid lesions were assessed microscopically.

Results

Biochemistry and Histology

Biochemical analyses revealed that the ibotenic acid lesions produced a 45% depletion of ChAT in frontal cortex and 44% depletion in sensorymotor cortex, while sparing levels in hippocampus in the NBM group as compared with the controls. The mean ChAT values (expressed as nmol/mg wet weight/hr $\pm$ Standard Error of the Mean, S.E.M.) for frontal, somatosensory, and hippocampal values, respectively, for the NBM group were 20.4 ($\pm$1.4), 5.9 ($\pm$1.0) and 10.5 ($\pm$0.6). Corresponding values for the control group were 37.2 ($\pm$5.7), 10.1 ($\pm$1.0), and 12.0 ($\pm$ 1.6). The NBM ChAT levels for both frontal and somatosensory cortex were statistically lower than their controls (t = 3.79 and 2.69, respectively, df = 22, p < .01). Hippocampal values for the two groups did not differ significantly.

Histological examinations revealed that the ibotenic acid injections destroyed at least some neurons in the area of the SI and LPO on both sides of the brain.

Activity

The frequency counts for each session were combined across the three days of testing to obtain an average frequency of activity for each behavior for each animal. In addition two new categories were created by combining the scores for Lying With Head Up and Lying With Head Down into the a category labeled Resting, and combining all other scores except Standing into a category labeled Activity. The scores for each behavior category for NBM and control groups at each test session and for combined dark and light test sessions were compared with the Student's t test.

The results revealed no consistent pattern of differences across test periods. At the 2100 h test period, the NBM group differed from the controls only on the frequency of standing (NBM mean 0.00, Control 0.15, t=2.78, df=30, p<.01), and the groups did not differ significantly on this measure at any other test period. At the 0600 test period, differences were revealed only on the measure of sniffing (NBM 1.31, Control 0.79, t=-2.28, df=30, p<.05). During the light sessions no significant differences in any of the behaviors were observed. No significant differences were observed at any of the test sessions for the derived measures of total Resting and total Activity.

For the combined data from dark and light test sessions, the NBM group differed from the controls only on the measure of Standing during the dark sessions (NBM 0.02, Controls 0.12, t=2.40, df=30, p<.05). Again no differences were revealed

between groups on the derived measures of total Resting or total Activity.

Discussion

The results of the present study do not confirm those of other investigations (5) showing increased locomotor activity induced by NBM lesions. The differences between the present study and others cannot be attributed to lack of depletion of cerebral ACh because the biochemical measures clearly show that the ibotenic acid lesions produced significant depletions of ChAT in frontal and somatosensory cortex, while sparing hippocampal ChAT.

The present study utilized a novel approach in measuring activity in an attempt to obtain measures of all of the most significant behaviors emitted by an animal. It is possible that this technique was not sensitive to the hyperactivity observed following NBM lesions in other studies. However, two subgroups of the lesioned and control animals in this study were tested for locomotor activity in an open field in a manner similar to that used in other activity studies, and the results also failed to reveal any indication of hyperactivity on the part of the NBM animals.

Another possible explanation could be that the lesions produced a transient hyperactivity that diminished by the time the behavioral observations in the current study were made. This is unlikely, however, because the hyperactivity response observed in other studies has been shown to be relatively permanent.

In light of the consistency of previous findings showing hyperactivity following lesions in the area of the NBM the most plausible interpretation of the present results is that the hyperactivity response can be dissociated from cortical depletions of ChAT. Previous studies in our labs (15) inducing ibotenic acid lesions in the area of the NBM using slightly different coordinates have replicated the hyperactivity finding reported by others (5) but have failed to obtain depletions of cortical ChAt. In addition, at least one other study involving lesions of the area of the NBM has reported a negative finding regarding motor activity (8).

The results of the present study suggest that depletion of cortical ChAT may not be a sufficient and necessary condition resulting in the hyperactivity syndrome reported following lesions in the area of the NBM. The specific neuroanatomical or neurochemical substrates for this behavioral syndrome remain to be revealed by studies designed to manipulate more specific areas of the ventral pallidum than have previously been attempted.

References

1. Arendt, T., Bigl, V., Arendt, A., & Tennstedt, A. (1983) Losss of neurons in the nucleus basalis of Meynert in Alzheimer's disease, paralysis agitans and Korsakoff's disease. Acta Neuropathologica, 61, 101-108.
2. Bendra, D. (1961). Components of general activity and the analysis of behavior. Psychological Review, 68, 205-215.
3. Bull, G., & Oderfeld-Nowak, B. (1971). Standardization of a radiochemical assay of choline acetyltransferase and a study

of the activation of the enzyme in rabbit brain. Journal of Neurochemistry, 18, 935-941.

4. Davies, P. (1979). Neurotransmitter-related enzymes in senile dementia of the Alzheimer type. Brain Research, 179, 319-327.
5. DuBois, B., Mayo, W., Agid, Y., Le Moal, M., & Simon, H. (1985). Profound disturbances of spontaneous and learned behaviors following lesions of the nucleus basalis magnocellularis in the rat. Brain Research, 388, 249-258.
6. Flicker, C., Dean, R. L., Watkins, D. L., Fisher, S. K., & Bartus, R. T. (1983). Behavioral and neurochemical effects following neurotoxic lesions of a major cholinergic input to the cerebral cortex in the rat. Pharmacology Biochemistry & Behavior, 18, 973-981.
7. Fonnum, F. (1975). A rapid radiochemical method for determination of choline acetyltransferase. Journal of Neurochemistry, 24, 407-409.
8. Hughey, D. J., Wenk, G. W., & Olton D. S. (1985). Lesions of the nucleus basalis magnocellularis impair maze learning in rats. Neuroscience Abstracts, 11, 833.
9. Jarrard, L. E. (1968). Behavior of hippocampal lesioned rats in home cage and novel situations. Physiology and Behavior, 3, 65-70.
10. Johnston, M. V., McKinney, M., & Coyle, J. T. (1979). Evidence for a cholinergic projection to neocortex from neurons in basal forebrain. Proceedings of the National Academy of Sciences, 76, 5392-5396.
11. Lehmann, J., Nagy, J. I., Atmadja, S., & Fibiger, H. C. (1980). The nucleus basalis magnocellularis: The origin of a cholinergic projection to the neocortex of the rat. Neuroscience, 5, 1161-1174.
12. Lerer, B., Warner, J., Zolcinski, M., Gamzu, E., & Friedman, E. (1983). Cortical cholinergic and behavioral impairment following kainic acid lesions of rat basal forebrain. Neuroscience Abstracts, 9, 97.
13. Perry, E. K., & Perry, R. H. (1980). The cholinergic system in Alzheimer's disease. In P. J. Roberts (Ed.), Biochemistry of dementia (pp. 135-183). Chichester: John Wiley.
14. Price, D. L. (1984). Neuropathology of Alzheimer's disease. In W. E. Kelly (Ed.), Alzheimer's disease and related disorders: Research and management (pp. 81-104). Springfield: Charles C. Thomas.
15. Sessions, G. R., Matthews, M. D., Swerdlow, N., Bahkit, C., & Koob, G. F. (1985) A comparison of cerebral acetylcholine or somatostatin depletion on avoidance conditioning, activity, and working memory in rats. Neuroscience Abstracts, 11, 1113.
16. Whitehouse, P. J., Price, D. J., Clark, A. W., Coyle, J. T., & DeLong, M. R. (1981). Alzheimer disease: Evidence for selective loss of cholinergic neurons in the nucleus basalis. Annals of Neurology, 10, 122-126.

Measuring Institutional and Occupational Values: Issues of Reliability and Validity

Michael J. O'Connell*
Air Force Institute of Technology

Abstract

Scales to measure the Moskos constructs of institutional and occupational orientation that were originally developed by Stahl and others (1978) are further analyzed to assess their reliability and validity. Reliabilities were found to be lower than is generally accepted, and analyses to demonstrate validity were only partially successful.

Background

Charles Moskos' arguments that the military is moving from an institutional to an occupational orientation has sparked interest by military sociologist and the Defense Department hierarchy. Moskos contends that focusing compensation and personnel policies on a market oriented philosophy tends to diminish traditional institutional values of service to country. His arguments have been seized by some critical of the all-volunteer force, by others defending traditional military benefits such as PXs, commissaries, and early retirement, and still by others concerned with a diminution of the military image, customs and courtesies. While some of the initial excitement Moskos generated has subsided, his ideas are still widely read and quoted.

Moskos has not gone unchallenged, and attempts to empirically measure some of his concepts have been only partially successful in validating his arguments. Cotton (1981) calls for conceptual clarity by arguing that the institutional-occupational concepts are opposite ends of the same continuum and that the term "vocational" is more appropriate than "institutional" to embody the notion of the military as a calling.

Stahl, Manley, and McNichols (1978) operationalized the institution-occupation concept with four survey questions to measure occupational attitudes and four to measure institutional orientations. The occupational questions measured attitudes toward the desirability of civilian employment and reactions to nonjob related aspects of the Air Force such as living on base. The institutional questions focused on impressions of the Air Force mission and national security and the need for discipline and close supervision. Because there was a high intercorrelation between the four questions in each set but not between questions in different sets, Stahl et al. (1978) concluded that the two concepts were basically different.

*The views expressed are the author's and do not necessarily reflect those of the U.S. Air Force or the Department of Defense.

The occupational scale developed by Stahl et al. (1978) has been criticized by Segal, Blair, Lengermann, and Thompson (1983) as measuring dissatisfaction rather than occupational orientation. Stahl et al. (1978) do not report reliabilities for their measures, but they do discuss validity by comparing average scores of various sub-groups. Senior military tended to be more institutional than junior military, and regular officers tended to be more institutional than reserve officers. In another study, Stahl et al. (1980) reported that Marine officers were less occupationally oriented than officers of other services but had institutional values comparable to officers in the other services. Analysis of Air Force personnel indicated that having a doctorate degree, being a physician, or having a high percentage of nonmilitary friends were all associated with lower institutional values and high occupational values.

Stahl's measures of the Moskos construct appear to have some validity, but some issues deserve further study. Stahl et al. (1980) show that it is possible for a person to be high on both institutional and occupational orientation, a phenomenon that Cotton's (1981) empirical approach would preclude and that Moskos' conceptual approach does not address. Stahl et al. argued that being high on both indicates that it is possible for a "pragmatic professional" to "coexist with traditional values and norms associated with the military." This study further investigates the empirical measurement of institutional or occupational orientation by focusing on a select group of officers. Both the reliability and validity of the Stahl et al. measures are assessed.

Methodology and Results

To gather additional occupational-institutional data U.S. Air Force Academy graduates were surveyed in April 1980. This survey was conducted by the Academy's Association of Graduates to collect attitude and opinion data on a variety of issues. The survey was mailed to 12,100 graduates, with usable returns from 5,914 for a total response rate of 49 percent. After deleting those not on active duty in 1980 and those who could not be matched with active personnel records, 3,347 active duty responses remained for this analysis. Because nonrespondents might be different from respondents, the results cannot be generalized to all Academy graduates.

The same items used by Stahl et al. were used to measure institutional and occupational values. As noted, some ambiguity exists as to whether these should be treated as two independent scales or as one scale. The correlation between two scales is -.176 for this data, a small but significant relationship. Means, standard deviations and reliability coefficients are shown in Table 1. The reliability of both scales is lower than what is generally accepted. For comparison, Cotton's (1980) six item Military Ethos Scale has a coefficient alpha of .78.

Table 1

Descriptive Statistics for Two Scales

	Mean	Standard Deviation	Coefficient Alpha
Occupational Scale	3.30	.67	.469
Institutional Scale	3.79	.86	.571

This research seeks to replicate some of the Stahl et al. construct validity results and also to assess how occupational and institutional orientation vary across other officer subgroups.

In this sample comparisons of institutional-occupational values can be made across officer grades. Based on Moskos' writing and Stahl's analysis we expect that as rank increases institutional values would increase and occupational values would decrease. Figure 1 displays average scores by grade. Except for second lieutenants, the expected results occur. Captains have the highest occupational orientation, and first lieutenants have the lowest institutional orientation, while colonels have the lowest occupational and highest institutional orientation. The values of second lieutenants are similar to those of majors or lieutenant colonels. Since this group of second lieutenants had recently graduated from the Academy, the scores for second lieutenants are probably best explained by the training and socialization processes of the Academy. Since all other officers were Academy graduates also, the data suggests that the effects of the Academy experience may dissipate over time particularly as the young officer nears the end of his or her obligated service. That senior officers are more institutional and less occupational may be explained by either continued service in the Air Force fostering such values or because the less institutionally and more occupationally oriented chose to leave active duty earlier.

An alternative grouping of officers by aeronautical rating allows another test of the validity of the measures. Because the primary mission of the Air Force is to fly, we might expect that pilots and navigators would have a stronger institutional orientation. By a similar logic nonflyers might be expected to have the strongest occupational orientation. Average institutional and occupational scores for subsets of the sample are shown in Figure 2. To control for the effect of grade, separate analyses by aeronautical rating were made for captains and for the combined group of majors and lieutenant colonels. Institutional scores for majors and lieutenant colonels are not significantly different across the four ratings. Pilot and navigator captains have lower institutional scores than other captains. Pilot captains have higher occupational scores than the other three groups. Only for majors and lieutenant colonels do pilots and navigators have lower occupational scores than the nonrated group. The data by aeronautical rating and grade do not suggest strong construct validity for the two measures.

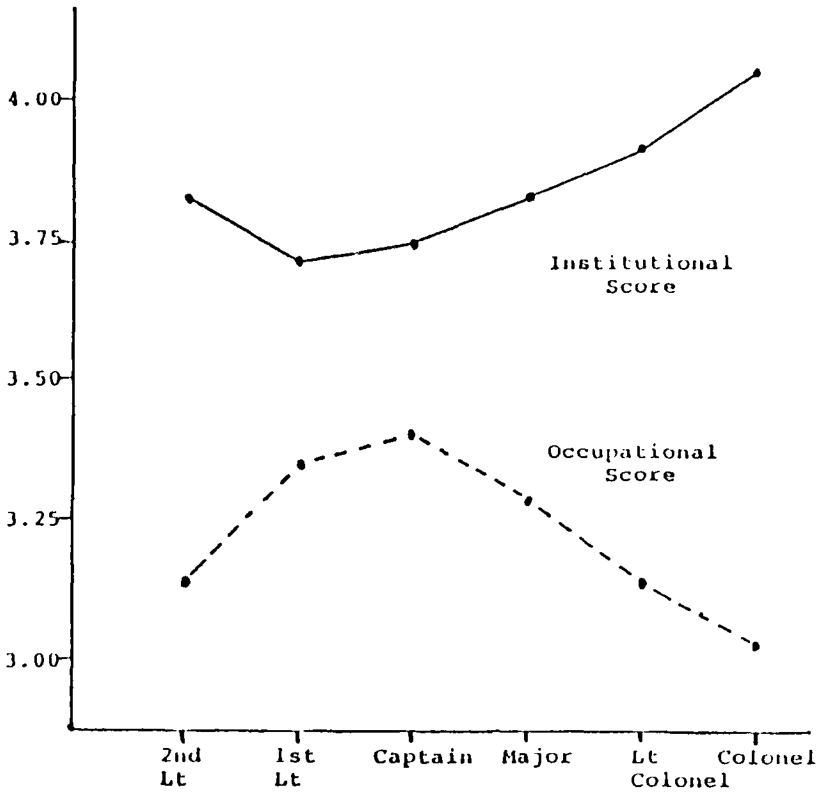


Figure 1

INSTITUTIONAL AND OCCUPATIONAL SCORES BY GRADE

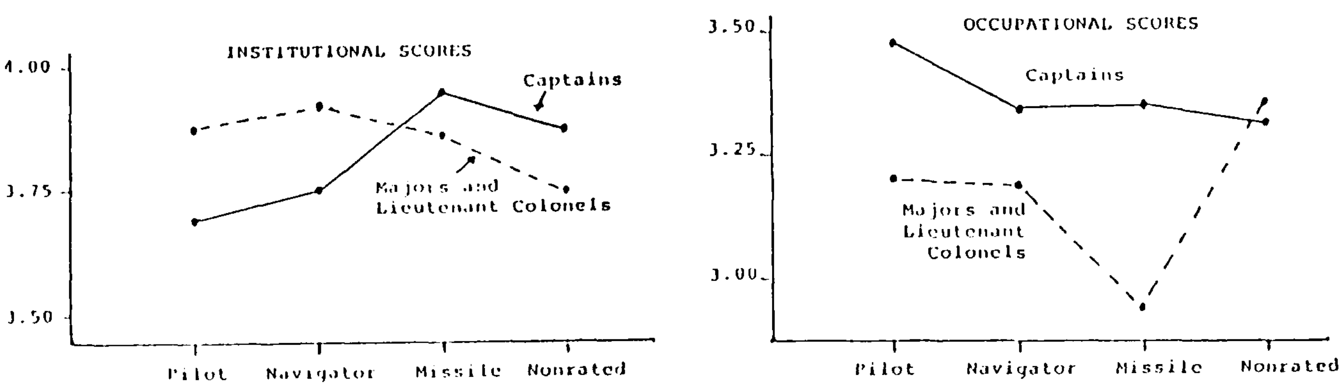


Figure 2

INSTITUTIONAL AND OCCUPATIONAL SCORES BY AERONAUTICAL RATING

Another way to assess the validity of the measures is to compare average values of high and average performers. We expect that high performers, as indicated by early promotion to one or more grades, would be more institutionally oriented and less occupationally oriented than those promoted on time. Those promoted early might be so oriented because the organization recognizes their institutional orientation and rewards it. Or those promoted early might become more institutionally oriented because they now are more committed to the institution and feel accepted. Lieutenant colonels and colonels promoted early two or more times tended to be more institutionally oriented and less occupationally oriented than those never promoted early or promoted early once. However, differences across performance levels were not statistically significant. The data from this analysis do not show strong construct validity for the measure, but this comparison is restricted insofar as those passed over for promotion are not included in the comparisons.

Overall, except for comparisons by grade, the preceding analyses indicate the Stahl et al. occupational and institutional measures may not have strong construct validity. Since the population is selective and homogeneous we may have hampered our ability to make as strong a case for construct validity as Stahl et al. did. In comparing subgroup differences for Academy graduates the measures have been subjected to a more demanding test than would be presented by a sample of a larger military population.

References

Cotton, C.A. Institutional and occupational values in Canada'a Army. Armed Forces and Society, 1981, 8, 99-110.

Segal, D.R., Blair, J.D., Lengermann, J.J., & Thompson, R.C. Institutional and occupational values in the U.S. military. In Changing U.S. Military Manpower Realities, F.D. Margiotta, J. Brown, & M.J. Collins (eds). Boulder, CO: Westview, 1983.

Stahl, M.J., Manley, T.R. & McNichols, C.M. Operationalizing the Mosko's institution-occupation model: An Application of Gouldner's cosmopolitan-local research. Journal of Applied Psychology, 1978, 63, 1978.

Stahl, M.J., McNichols, C.W., & Manley, T.R. An Empirical Examination of the Moskos institution-occupation model. Armed Forces and Society, 1980, 6, 257-269.

Stahl, M.J., McNichols, C.W., Manley, J.R. A longitudinal test of the Moskos institution-occupation model: A three-year increase in occupational scores. Journal of Political and Military Sociology, 1981, 9, 43-47.

Initial Examination of the Construct Validity of the Organizational Assessment Survey*

Lawrence O. Short, Lt Colonel, USAF
Air Force Human Resources Laboratory (AFHRL)

James K. Lowe, Captain, USAF
1605th Civil Engineering Squadron

Janice M. Hightower, Captain, USAF
Detachment 5, Air Force Operational Test and Evaluation Center

Abstract

This study concerns initial standardization of the Organizational Assessment Survey (OAS) with emphasis on item sensitivity and construct validity of the revised instrument. Data were derived from 1,583 subjects from three USAF bases. Construct validity results were encouraging, as theoretical and cross instrument relationships were largely verified.

Introduction

In 1981, the Directorate of Research and Analysis of the Leadership and Management Development Center (LMDC) began work on a revision of the primary data gathering instrument used in the LMDC consulting process, the Organizational Assessment Package (OAP). The purpose of this report is to discuss the results of initial standardization research performed on the second generation OAP, now called the Organizational Assessment Survey (OAS). A previous work (Short, Lowe, & Hightower, 1985) dealt with deriving the OAS factor structure and testing each factor's internal consistency reliability. The present report considers OAS item sensitivity and construct validity.

In its present form, the OAP survey consists of a computer-scored response sheet and a 109-item (93 attitudinal and 16 demographic) booklet. Responses use a scale of one to seven, with a value of "1" generally indicating strong disagreement or dissatisfaction with the question or statement, and a "7" indicating strong agreement or satisfaction. Through factor analysis, the 93 attitudinal items are combined into factors representing the job itself, supervisory satisfaction, work group effectiveness, and organizational climate issues.

Administration of the survey is the first step in the consultation process. The survey is given to a stratified random sample of the organization to which LMDC has been invited. The results of the survey are an important feature in the assessment of task, supervision, climate, and productivity in an organization. The results are handled in a confidential manner between LMDC and the client. After approximately five to six weeks for analysis, consultants return to the organization to provide feedback of data to commanders and supervisors.

*All work was completed while authors were members of the Leadership and Management Development Center, Research and Analysis Directorate.

When organizational problems are encountered, a consultant and supervisor develop a management action plan designed to resolve the problem at that level of the organization. Within six to nine months, the consulting team returns to readminister the survey instrument as a means to help assess the impact of the consulting process.

The data from each OAP administration effort are stored in a cumulative data base currently containing over 200,000 records for research purposes. These data are aggregated by work group codes developed for this instrument. The data may be recalled by demographics such as personnel category, age, sex, Air Force Specialty Code, pay grade, time in service, and educational level.

The OAP was developed jointly by LMDC and AFHRL at Brooks AFB, Texas (Hendrix & Halverson, 1979a; 1979b). More recently, additional standardization work has been done with the OAP. Short and Hamilton (1981) provided evidence of the factor-by-factor reliability of the instrument considering both internal consistency and test-retest (stability) aspects. In addition, Short and Wilkerson (1981) provided evidence in support of the group differences aspect of OAP construct validity. Webster (1982) also studied construct validity of the leadership and organizational climate areas of the OAP by using a modified multi-trait, multi-method approach, favorably comparing the OAP to the Survey of Organizations (Taylor & Bowers, 1972). Finally, the stability of the OAP factor structure was studied across selected functional area and demographic groups (Hightower & Short, 1982) and across intervals of time (Hightower & Short, 1983). These studies yielded a slightly different factor structure than that currently in use, but showed the revised structure to be extremely stable across all comparison groups. These studies combined with several years of experience using the instrument in the consulting process pointed out ways the OAP could be revised to enhance the process and the accuracy and precision of organizational diagnoses.

Method

Overview and Subjects

The subjects of the OAP-OAS data gathering came from three operational bases in the continental United States. Since the surveys were administered by LMDC consultants, the test bases were preselected via the consultant's schedule. The data were gathered by LMDC consultants both as part of the consulting process and as an effort to test the new OAS instrument. The subjects responded to both the OAP and the OAS and were then provided the opportunity to verbally express their thoughts concerning the face validity of the OAS instrument. The responses from each person on the two surveys were linked using special coding.

A total of 1,583 personnel took the back to back OAP-OAS surveys. The sample consisted of 85.2% males and 14.8% females, compared to the 1984 Air Force ratio of 89% male to 11% female. Ages ranged from 17 to 63, although 96% were younger than 47. The sample consisted of 13% officers, 79% enlisted, 7% General Schedule (GS) civilians, and 1% Wage Board (WB) civilians. The Air Force officer/enlisted ratio is 17% officer, 83% enlisted.

Note that if the civilians were removed, the officer/enlisted ratio of the OAS sample would be 14% officer to 86% enlisted. Fifty-four percent (54%) had made either one or two PCS moves, and 24% had been on at least one unaccompanied PCS tour. Over 56% of the personnel had daytime work schedules. Racially, 76% of the sample were Whites, 13% were Blacks (compared to the Air Force's 15%), and 14% were others. Over 27% had more than 12 years in the Air Force, while 21% had less than two years of total service time.

Procedure

The version of the OAS used for the present study contained 104 attitudinal items for test purposes and 10 demographic items in addition to those in the OAP. During each survey administration, the OAP was administered prior to the OAS to prevent contamination of OAP results. After completing both surveys, participants were asked to complete a short questionnaire about the OAS. In addition, 384 respondents were randomly selected and verbally polled to see if they had additional comments about the survey. If so, those comments were recorded on the questionnaire.

One of the issues that must be addressed in survey work is the adequacy of items and factors as dependent variables. This adequacy deals with both sensitivity and validity of the measures. If they are adequate, we should be able to predict relationships among and between attitudinal and demographic variables. Carmines and Zeller (1979) list three distinct steps toward demonstrating construct validity: specifying the theoretical relationships between concepts, examining empirical relationships between concepts, and interpreting empirical evidence. Thus, empirically verifying predicted theoretical relationships with survey results would provide important evidence about both survey construct validity and the sensitivity of the resulting measures.

For purposes of this study, expectations were stated in four areas for each of three data bases. The three data bases consisted of (1) the existing cumulative OAP pre-consultation data, in addition to (2) the OAP results and (3) the OAS results which occurred when the two surveys were administered together in this study. Each expectation was tested on each data base.

The expectations tested were:

1. There will be a significant positive correlation between measures of task significance and pride;

2. Perceptions of task autonomy will be significantly lower for personnel in the maintenance career field than for personnel in other career fields;

3. Perceived co-worker performance will be significatly positively correlated with career intent; and

4. Training will be significantly positively correlated with work group effectiveness.

The reason for this effort centers on consistency of results. The predictions tested were drawn from consultants' observations of what actually happens in the field as consulting interventions take place. It was important, therefore, to again test within the entire pre-consultation data base to see if these observations were reflected in this data.

Beyond this, however, is the need to establish consistency between OAP results and OAS results when administered in the same environment. The issue centers on whether or not similar results are obtained in each case. If so, we may be more confident that the OAS is measuring what it is supposed to measure and that items are sufficiently sensitive for use as dependent variables. We may also be confident that the OAS operates in a way very similar to the OAP and thus be able to generalize a larger amount of OAP-specific research to the new instrument. All expectations were tested using Pearson product-moment correlation coefficients ($p < .05$) or independent groups *t*-tests ($p < .05$).

Results

The first relationship tested was the expectation that task significance and pride would have a significant positive correlation. This relationship was tested by finding the Pearson product-moment correlations between the Task Significance and Pride factors of the OAP and two comparable items on the OAS. For the pre-consultation data the correlation was .47 (*n*=87,873). The correlations were .48 (*n*=1,528) and .47 (*n*=1,566) for OAP and OAS surveys in this study. All were statistically significant.

The second expectation was that task autonomy would be significantly lower for maintenance personnel than for personnel in the remainder of the data base exclusive of the maintenance career field. The results of the *t*-tests using the Task Autonomy factor of both the OAP and the OAS supported our expectation for the pre-consultation OAP data base, the OAP data base and the OAS data base. The *t*-statistics were 54.30 (*n*=80,819), 6.16 (*n*=1,314) and 4.98 (*n*=1,531), respectively, which were all statistically significant in the expected direction.

The third expectation was that perceived co-worker performance would be positively correlated with career intent. While not as strong as expected, both the OAP and the OAS displayed positive correlations and were highly consistent in their results. Both the pre-consultation OAP data and the OAS data resulted in a Pearson product-moment correlation value of .21 (pre-consultation *n*=77,550; OAS *n*=1,380) between career intent and perceived co-worker performance items. The OAP data within the OAP-OAS data base had a lower correlation, .16 (*n*=1,417). Again, all correlations were statistically significant.

The final expectation was that training would be positively correlated with work group effectiveness. OAS data showed a Pearson product-moment correlation value of .48 (*n*=1,273) between Work Group Effectiveness and Training factors. The OAP data and pre-consultation data base also showed statistically significant positive correlations between the Work Group Effectiveness and Training factors (*r*=.44, *n*= 1,370, and *r*=.38; *n*=77,690, respectively).

A Final Comment

The results of the expectations tested, then, indicated strong similarities between the OAP and OAS instruments. The items generally behaved as expected, and the OAS item outcomes were consistent with OAP item outcomes. The OAS, therefore, seems appropriate to replace the OAP. Items appear to be sensitive enough with sufficient instrument construct validity to support use of the OAS as either a consulting or research instrument.

References

Carmines, E. G., & Zeller, R. A. (1979). Reliability and validity assessment. Beverly Hills, CA: Sage.

Hendrix, W. H., & Halverson, V. B. (1979a). Organizational survey assessment package for Air Force organizations (AFHRL-TR-78-93). Brooks AFB, TX: Air Force Human Resources Laboratory.

Hendrix, W. H., & Halverson, V. B. (1979b). Situational factor identification in Air Force organizations (AFHRL-TR-79-10). Brooks AFB, TX: Air Force Human Resources Laboratory.

Hightower, J. M., & Short, L. O. (1982). Stability of the Organizational Assessment Package factorial validity across groups. Paper presented at the 90th Annual Convention of the American Psychological Association, Washington, DC.

Hightower, J. M., & Short L. O. (1983). Temporal stability of the Organizational Assessment Package factor structure. Paper presented at the 91st Annual Convention of the American Psychological Association, Anaheim, CA.

Short, L. O., & Hamilton, K. L. (1981). An examination of the reliability of the Organizational Assessment Package (LMDC-TR-81-2). Maxwell AFB, AL: Leadership and Management Development Center.

Short, L.O., Lowe, J.K., & Hightower, J.M. (1985). Initial standardization of an Air Force organizational assessment survey instrument. Proceedings of the 27th Annual Conference of the Military Testing Association. San Diego, CA: Military Testing Association.

Short, L. O., & Wilkerson, D. A. (1981). An examination of the group differences aspect of the construct validity of the Organizational Assessment Package. Proceedings of the 23d Annual Conference of the Military Testing Association. Washington, DC: Military Testing Association.

Taylor, J. C., & Bowers, D. G. (1972). Survey of Organizations. Ann Arbor: Institute for Social Research, University of Michigan.

Webster, A. H. (1982). An assessment of the construct validity of the Leadership and Management Development Center's Organizational Assessment Package (Report number 82-2640). Maxwell AFB, AL: Air Command and Staff College.

Federal Service Academies: Is the "Total Institution" Analogy Appropriate?

Fred M. Rosa, Jr.
Gwendoyln R. Stevens

Department of Humanities
United States Coast Guard Academy

Abstract

To evaluate the usefulness of the total institution concept (Goffman, 1961) as a model for socialization at military service academies, 32 first-year and 36 fourth-year cadets at the U. S. Coast Guard Academy were administered the Survey of Interpersonal Values (Gordon, 1960) and the Survey of Personal Values (Gordon, 1967). Data analysis revealed that for 6 of the 12 sub-scales (specifically, conformity, recognition, independence, leadership, goal orientation, and variety), the responses for first and fourth-year cadets differed beyond chance. The extent to which these results provide support for the total institution construct is discussed along with an alternative explanation of the data, i.e., rather than being resocialization outcomes in any meaningful sense, the value changes observed are instead situational cadet responses to environmental pressures which are more analogous to rites of passage than total institution resocialization mechanisms.

Introduction

In examining socialization processes and outcomes at federal service academies, social scientists associated with the various disciplines often utilize Goffman's (1961) total institution concept. According to Goffman, the total institution is a "place of residence and work where a large number of like-situated individuals, cut off from the wider society for an appreciable period of time, together lead an enclosed, formally administered round of life" (1960: xiii). The total institution effects a "breakdown of the barriers ordinarily separating these three spheres of life [work, sleep and play]" (1960: 6) through its four characteristic parameters: (1) life's activities are confined to a single physical location under the total control of a single authority; (2) all activities take place in the company of and together with a large group of similarly-situated individuals; (3) all activities are highly regimented and tightly scheduled; and (4) all activities are conducted in accordance with a deliberate plan the purpose of which is to further the goals of the institution.

While Goffman's primary focus is on mental hospitals and prison settings, "army barracks" and "boarding schools" are included among his many examples of total institutions. More importantly for present purposes, numerous subsequent researchers utilize the total institution concept in their exploration of socialization phenomena at military academies and training centers. [For examples, see Rootman (1972), Zurcher (1972), and Arbin and Dobrofsky (1978).]

While these psychologists and sociologists presumably select the total institution model in part because military academies and training centers so clearly fit the descriptive criteria, an additional consideration is their underlying assumption (at least implicitly) that total institutions are optimal environments for resocialiation. In other words, the total institution concept is utilized because researchers assume that, in addition to being accurately descriptive of critical environmental parameters, the concept will assist them in their study of resocialization outcomes or behavioral/value profile changes. Recent evidence, however, suggests that the total institution model may be inappropriate or at least have significant limitations for application in the federal service academy setting.

In the first place, the federal service academy, while very similar to the total institution in many fundamental aspects, is at once strikingly different from Goffman's asylum. Perhaps the three most important distinctions are: (1) the unique institutional purpose (affording future military officers both an academic degree and professional training); (2) the self-selection of military academies by young applicants (thus promoting a pre-existing applicant-institutional value match); and (3) the substantially greater freedom cadets and midshipmen enjoy (even to the point of voluntary resignation).

A second consideration is that, despite the operation at military academies of the full array of resocializing mechanisms which typify the total institution (Goffman, 1961), rather surprisingly there is no compelling evidence that appreciable value resocialization takes place, i. e., that value profiles change to any measurable extent as cadets/midshipment adopt new roles and associated behaviors over their four-year academy tenures. In fact, several recent studies at the U. S. Military Academy (Priest, Fullerton and Bridges, 1982) and the U. S. Coast Guard Academy (Rosa and Stevens, 1984) indicate that there may be little value change indeed. The West Point study, which was particularly exhaustive, shows no important changes in independence, creativity, religiousity, honesty, kindness, drive to achieve and other relevant values. The few changes reported were only of moderate size, and included an increase in political conservatism and a decrease in commitment to the outside world as a reference group. In evaluating the foregoing data, however, it should be noted that the military academy's resocialization mechanisms may work primarily to guide attrition, effectively weeding out those cadets and midshipmen whose value profiles are not consistent with those of the military (Rootman, 1972).

The present study sought to further test the applicability of the total institution analogy to the military academy environment by comparing the composite value profiles of first and fourth-year cadets at the United States Coast Guard Academy. Is there significant change in cadet value profiles, and, if so, what is the nature and direction of value change?

Method

The subjects for the present cross-sectional study were first and fourth-year cadets at the U. S. Coast Guard Academy in New London, Connecticut. While by far the smallest of the four military service academies, Coast Guard is very similar to its sister academies in virtually every other respect, including institutional purpose and orientation, academic and professional training curricula, and cadet demographics.

The research instrument consisted of two standard tests - the Survey of Interpersonal Values (Gordon, 1960) and the Survey of Personal Values (Gordon, 1967) - both of which approach value assessment by directly asking respondents what they consider most important. The surveys afforded cadet subjects a total of 60 forced-choices among a triad of value statements; the responses yielded independent ipsative measures of 6 interpersonal values (support, conformity, recognition, independence, benevolence and leadership), and 6 personal values (practical-mindedness, achievement, variety, decisiveness, orderliness and goal orientation). Thus the research instrument provided both interpersonal and personal value profiles for each individual cadet.

The procedure adopted involved administering the two-survey instrument to first and fourth-year cadets, and then effecting a cross-sectional comparison of the composite value profiles for both years. All first-year cadets received the instrument during the regular "swab" summer testing program several weeks after reporting to the Academy; all fourth-year cadets received both an initial and a follow-up request for their voluntary participation during the fall semester of their senior year. Thirty-six fourth-year cadets returned their surveys (subsequently four forms had to be eliminated), and thirty-six randomly-selected first-year forms were therefore selected to effect the comparison. In terms of gender breakdown, there were 30 males and 6 females among the first-year cadets, and 27 males and 5 females among the fourth-year group.

Results

Responses to the two surveys were analyzed using an ANOVA procedure from SPSSX, a statistics software package. Initial analysis of the 12 sub-scales included both gender and year group as variables. Only the SIV benevolence sub-scale revealed a difference beyond chance for gender ($F = 4.99$, $df = 1$, $p < .03$), with the women's mean being higher ($\bar{x} = 20.55$) than that for men ($\bar{x} = 16.19$).

All of the data were then collapsed for gender and reanalyzed only for the variable of year group. Table 1 displays the results of the analysis: in summary, there was a difference beyond chance for four SIV sub-scales (conformity, $F = 20.92$, $df = 1$, $p < .001$; recognition, $F = 4.01$, $df = 1$, $p < .05$; independence, $F = 4.24$, $df = 1$, $p < .05$; and leadership, $F = 4.22$, $df = 1$, $p < .04$), and for two SPV sub-scales (variety, $F = 7.40$, $df = 1$, $p < .008$; and goal-orientation, $F = 4.79$, $df = 1$, $p < .03$).

Table 1

Personal and Interpersonal Value Sub-scales by Year Group

Personal Values	1st Year Cadets		4th Year Cadets		Comparison
	$\bar{X}$	*SD*	$\bar{X}$	*SD*	*PROB (ANOVA)*
Practical-Mindedness	11.78	4.28	11.45	3.98	ns
Achievement	19.22	5.38	18.07	4.57	ns
Variety	12.22	7.18	17.58	8.50	.008
Decisiveness	16.89	4.68	15.83	5.14	ns
Orderliness	11.93	6.04	15.53	7.95	ns
Goal Orientation	17.86	5.07	15.06	5.58	.03

Interpersonal Values	1st Year Cadets		4th Year Cadets		Comparison
	$\bar{X}$	*SD*	$\bar{X}$	*SD*	*PROB (ANOVA)*
Support	16.06	4.89	15.50	5.24	ns
Conformity	14.78	5.46	8.78	5.32	.001
Recognition	10.94	4.70	13.16	4.36	.05
Independence	13.92	6.96	17.66	8.02	.05
Benevolence	17.83	4.74	15.84	7.30	ns
Leadership	16.28	5.88	19.17	5.35	.04

Discussion

The predicted hypothesis - consistent with the researchers' stated intention of testing the total institution construct - was that there would be no difference between the composite value profiles for first and fourth-year cadets. The results did not support the hypothesis (i. e., 6 of 12 value sub-scales yielded significant differences), and in fact provide some evidence for the applicability of the total institution concept.

However, while the data fail to support the hypothesis, there is reason not to reject the basis for its formulation. While the new first-year cadets were administered surveys within several weeks of their initial arrival at the academy, it may be that even this early time-frame is too late to serve as an adequate test of the value impact of the institutional experience. Perhaps each successive class arrives at the academy with a given value profile (which the authors at least have yet to survey before the "swabs" are totally caught up in their new world), modifies its value orientation in response to environmental pressures (which are strongest early in cadet experience and precipitate changing values temporarily to render them consistent with required behaviors), and then reclaims its initial value orientation as the intensity of the academy experience tapers off and graduation appears certain.

If these dynamics exist, then the conformity and goal orientation results (the two sub-scales where first-year cadets scored higher), might well reflect the strong environmental pressures for conformity and the total commitment which new cadets must make in order to survive the summer. Similarly, the sub-scale comparisons for leadership, independence, recognition and variety (the four on which fourth-year cadets scored higher) might reflect the comparatively less-structured and more autonomous nature of a cadet's final year at the academy. Stated differently, the fourth-year cadet has left behind his/her oppressive swab summer world in which opportunities for leadership were minimal, independence was not reinforced, a low-profile was an important coping asset, and variety was eclipsed by the narrow scope and uniformity of daily routine.

Experimental testing of the total institution concept as applied to military service academies is fairly new, and therefore little direct comparison with previous research is possible. Using a longitudinal approach, Priest, Fullerton and Bridges (1982) also report selected changes in cadet personality and value formation between the first and fourth years at West Point. Of the comparable sub-scales, they note decreases in academic achievement, leadership and independence. Yet their overall conclusion, which takes all of the personality dimensions and value scales assessed into account, is that West Point cadets do not change significantly over their four years.

Summary and Conclusion

On balance, then, the applicablilty of the total institution concept in the military academy context appears to require further examination. A military academy definitely fits the descriptive criteria Goffman advanced for the total institution, and there is little doubt that the military academy experience is an intense one for the individual cadet or midshipman. Moreover, there is evidence - including that reported in the present study - of some value change over the four-year time frame. In the authors' view, however, what evidence there is does not come close to conclusively validating the concept's applicability.

Additional and more comprehensive research is certainly indicated. Incoming cadets/midshipmen should be tested much earlier (ideally on the very day of their arrival), and the use of different value instruments should be considered (in order to ascertain whether resocialization mechanisms are impacting values other than those surveyed to date). The authors also submit the necessity of evaluating the results of socialization outcome studies in the context of a military academy's fundamentals goals, and would welcome the opportunity to coordinate research with the other federal service academies.

References

Arkin, W. and Dobrofsky, L. Military Socialization and Masculinity. Journal of Social Issues, 1978, 34(1), 151-168.

Brim, O. and Wheeler, S. Socialization After Childhood: Two Essays. New York: Wiley, 1967.

Dornbusch, S. The Military Academy as an Assimilating Institution. Social Forces, 1955, 33, 316-321.

Goffman, E. Asylums. Garden City, New York: Doubleday, 1961.

Gordon, L. Survey of Interpersonal Values: A Manual. Chicago: Science Research Associates, 1976.

Gordon, L. Survey of Personal Values: A Manual. Chicago: Science Research Associates, 1984.

Mortimer, J. and Simmons, R. Adult Socialization. Annual Review of Sociology, 1978, 4, 421-454.

Ondrack, D. Socialization in Professional Schools: A Comparative Study. Administrative Science Quarterly, 1975 (March), 20, 97-103.

Pascale, R. The Paradox of "Corporate Culture": Reconciling Ourselves to Socialization. California Management Review, 1985 (Winter), 27(2), 26-41.

Priest, R., Fullerton, T. and Bridges, C. Personality and Value Changes in West Point Cadets. Armed Forces and Society, 1982 (Summer), 8(4), 629-642.

Rokeach, M. The Nature of Human Values. New York: Free Press, 1973.

Rootman, I. Voluntary Withdrawal From A Total Adult Socializing Organization. Sociology of Education, 1972, 45, 258-270.

Rosa, F. and Stevens, G. Environmental Structure and Resocialization at the U. S. Coast Guard Academy. Presented at the 92nd Annual Meeting of the American Psychological Association. Toronto, Ontario, Canada. 24-28 August 1984.

Schein, E. H. The Individual, the Organization, and the Career: A Conceptual Scheme. Journal of Applied Behavioral Science, 1971, 7, 401-426.

Van Maanen, J. People Processing: Strategies of Organizational Socialization. Organizational Dynamics, 1978 (Summer), 19-36.

Zurcher, L. The Naval Recruit Training Center: A Study of Role Assimilation in a Total Institution. In The Social Dimensions of Work, ed. C. Bryant. Englewood Cliffs, New Jersey: Prentice Hall, 1972, 298-312.

Authors' Note: The views discussed in this report are those of the authors and not the official position of the U. S. Coast Guard Academy, the U. S. Coast Guard, or any other government agency.

NAVY FLAG OFFICER EQUAL OPPORTUNITY PRETRAINING ASSESSMENT SURVEY

Lt Col Frank C. Gentner
Defense Equal Opportunity Management Institute

Abstract

This Survey was developed by the Defense Equal Opportunity Management Institute (DEOMI) in cooperation with Navy Military Personnel Command's Equal Opportunity Office (NMPC-61) to identify flag officer opinions on equal opportunity (EO) issues and the suitability of the proposed course objectives for a Flag Officer EO Management Seminar. Requested by the Deputy Chief, Naval Operations for Manpower, Personnel and Training, the 50-question survey sampled 252 of the 324 flag officers sent surveys. Flag officers saw a direct link between EO and accomplishing the Navy mission (97%). They (74%) endorsed the need for an EO program, although 15 percent believed the program should be eliminated. When asked about the effect of the Navy EO program, they believed that the Navy's EO program was only moderately effective in deterring racism/sexism. Responding to open-ended questions, over 60 percent enumerated both strengths and weaknesses of the current EO program. They commented, although the EO program has done much to promote harmony through awareness, it is now time to move the program fully into the chain of command. By far the most significant elements of an effective EO program, in the opinion of flag officers, are the commanding officer's role in modeling active support of EO, and resolving EO issues through the chain of command. The most significant EO issue of the 1980s and 1990s was perceived to be coping with the expansion of women into nontraditional work roles. In addition, flag officers endorsed each proposed seminar objective by 75 to 94 percent; they also suggested additional topics.

BACKGROUND

At the request of the Deputy Chief, Naval Operations for Manpower, Personnel and Training (Chief of Navy Personnel), DEOMI developed objectives for a proposed Equal Opportunity Management Seminar for flag officers. To verify that the objectives were necessary and sufficient, DEOMI's Evaluation Division developed a survey containing 67 items to solicit the opinions of flag officers toward equal opportunity issues and the proposed seminar objectives. During staff coordination, the Navy Military Personnel Command's Equal Opportunity Office (NMPC-61) modified several items and shortened the survey to 50 items, plus the five open-ended questions proposed by DEOMI. The survey was approved by Vice Admiral Lawrence and forwarded to 324 Navy flag officers.

METHODOLOGY

The survey contained fifty degree-of-agreement questions using a Likert type 5-point response scale. Eleven of the items contain a midpoint of "no opinion" with the extremes being "strongly disagree" and "strongly agree." Other standardized items ranged from "1" (no degree) to "5" (extreme degree) or (completely). Five open-ended questions solicited flag officer comments about the strengths and weaknesses of the Navy's EO program; Navy EO challenges of the 1980s and 1990s; comments on or suggestions for new seminar objectives; and general comments on the seminar, the EO issues, or the Navy's EO program. Scaled items were analyzed by groupings in the following areas:

- EO's link with readiness
- Need for an EO program
- Intergroup relations needing attention
- Strengths and weaknesses of Navy's EO program
- Elements important to an effective EO program
- EO issues of the 1980s and 1990s
- Degree Navy deals with each type of discrimination
- Degree Navy assures equity
- Degree EO program understood/effective
- Degree military education, training, and experience prepared flag officers to provide equal opportunity and treatment to subordinates
- Degree proposed objectives for Flag Officer EO Management Seminar on target

Responses were analyzed in terms of mean rating, percent responding, and relative position within each topic area. Narrative comments were clustered within each topic and sorted by similarity. Narrative comments were used to validate items, since prevalidation was impossible due to time constraints. Computer display programs for analysis were developed by the DEOMI Evaluation Division's Chief Programmer using the Statistical Package for the Social Sciences (SPSS) statistical algorithms and report writer programs. The Chief of DEOMI's Evaluation Division developed the survey, analyzed the data and wrote the report. The survey was mailed by NMPC-61 on 14 Mar 1985 and was officially closed out on 7 May 1985.

Flag officers were asked to provide their type of assignment, major command, and race/ethnic group. No name or social security number was required. Returns were tracked by a survey control number placed on the return envelope, but 23 removed their control number before returning their booklets. Thus, it appears that the flag officers enjoyed a great deal of anonymity in answering the survey. From their comments, it seems that they did not take "the party line" for appearance sake. If they disagreed with aspects of the equal opportunity program, they stated their objections. Thus, it is likely that the survey is relatively free from this type of bias.

SAMPLE

The 252 responding flag officers comprise 77.8 percent of the 324 flag officers who were sent surveys. The largest percentages of respondents served as CNO staff officers (23%), major command staff officers (18%), major shore installation commanders (16%), and those who categorized themselves as "other" (25%). Largest participation from major commands came from CNO (33%), CINCPACFLT (13%), and CINCLANTFLT (12%), with 17 percent falling into an "other" category.

Eighty-two percent of survey respondents categorized themselves as Caucasian. Two flag officers identified themselves as Black; three, Hispanic; and three, Asian American or Pacific Islander; and two American Indian or Alaskan Native for a total minority representation of approximately four percent. A total of 14 percent categorized themselves as "other", left the item blank, or gave a multiple response. Only one of the three women sent a survey replied, unless the other two women could be among the 23 unknown whose booklets were missing survey control numbers. Of the 252 respondents, 132 (52%) were commodores; 81 (32%) were rear admirals; 17 (7%) were vice admirals; and 23 (9%) were unknown. While 89 percent of the booklets were sent to Continental United States (CONUS) locations, only 62 percent of the returns could be tracked as returned from CONUS locations. The overseas return rate was closer to the percentage sent. The locations of seven percent of respondents were not identifiable. Fifty-one percent of respondents indicated they were familiar with the Navy's new Command Managed Equal Opportunity (CMEO) program. The survey had no way of identifying whether the flag officer to whom it was sent actually completed the survey, or whether some lower-ranking person filled it out instead of the flag officer. The assumption is that most flag officers honored Vice Admiral Lawrence's request to "take a few minutes to complete" the survey.

FINDINGS

EO and Readiness

Based on their experience, 98-99 percent of flag officers surveyed saw a strong link between EO and readiness. Also, 97 percent believed EO is important to Navy mission accomplishment.

Need for an EO Program

When asked if they agreed that the EO program has served its usefulness and should be eliminated, 15 percent agreed, 11 percent had no opinion, and 74 percent disagreed. Thus, three-fourths agree that the EO program is necessary. Conversely, 15-26 percent did not support maintaining an EO program, despite the fact that 97 percent believe EO is important to accomplishing the Navy mission.

Intergroup Relations Needing Attention in Navy

When asked the degree to which intergroup relations need attention in the Navy, only 17-38 percent thought the need for attention to any group was high, while 68 to 89 percent of flag officers thought these intergroup relations needed a "moderate" or greater amount of attention. Eighty-nine percent believed that Black-majority relations needed attention from a "moderate" to high degree. Eighty-five percent believed male-female relations needed attention from a "moderate" to high degree. In both these areas, 38-40 percent believed the need for attention was "great" or "extreme." Significantly lower in emphasis were Hispanic-majority relations, with only 27 percent indicating these relations need a "great" or "extreme" amount of attention. Flag officers were evenly divided on the degree of attention needed for "other" majority relations. Flag officers also indicated some concern about inter-minority group relationships, with less emphasis on Hispanic-"other" relationships. Thus, flag officers do believe that minority-majority, male-female, and to a lesser degree minority-minority group relationships need attention in the Navy.

Strengths and Weaknesses of Navy's EO Program

When asked to discuss the strengths of the EO program, 161 of the 252 respondents (64%) took time to comment on various strengths. Write-in comments focused on these areas:

° Fairness and equity afforded to all
° Improved communication, awareness, and sensitivity
° The fact a program exists--as security, a reminder, actively solving problems

- Command support/commitment, visibility of support, and use of the chain of command
- Education and publicity
- Increased teamwork which improves readiness

One hundred and fifty-seven (62%) commented on weaknesses of the Navy's EO program. They focused on the following areas:

- Failure to use the chain of command
- EO program management, directives, and procedures are too bureaucratic (although this area seemed to focus more on civilian EEO procedures than military EO)
- Lack of timely, accurate climate assessment and issue reporting
- Inequitable handling of EO issues/reverse discrimination/manipulation of the program
- Selection/qualifications/training/competence of personnel assigned to work EO issues
- Amount of emphasis placed on EO by various commanding officers--from inaction to over-reaction
- Gender/racial problems still exist which need resolution or improvement, particularly the utilization of women in the Navy
- Need for widespread understanding/awareness and visibility of equal opportunity principles--that EO is everybody's business.

Elements Important to an Effective EO Program

When asked the degree to which various EO program elements were important to an effective program, they saw the commanding officer's role in modeling active support of equal opportunity, and resolving EO issues through leadership in the chain of command as most important, with 89-93 percent strongly agreeing. Next most important (but significantly lower) was developing command programs sensitive to local EO issues (67 percent rating this item high). Also seen as quite important was climate assessment to determine problem areas. Moderately important program elements were EO education and affirmative action. Only 75 percent saw publicity of EO efforts as moderately or more important to an effective EO program. All major elements of an EO program were seen as at least "moderately" important but commanding officer modeling support and resolving issues through the chain of command were, by far, the most important program elements.

EO Issues of 1980s and 1990s

When asked the degree EO issues will be challenges to the Navy in the 1980s and 1990s, the most challenging issue was expansion of women into nontraditional roles. Twenty flag officer comments echoed this concern and gave illustrations of problems likely to be encountered. Significantly lower, but next highest were:

- Upward mobility of minorities and women
- Accession of qualified minorities
- Off-base treatment of minorities in the local community

Unqualified minorities or women promoted past their level of competence, retention of minorities and women, the Navy's public image regarding EO, and unequal minority/female distribution across NECs, were seen as "moderate" challenges. Surprisingly, preventing sexual harassment and racial/ethnic tension were rated as only "moderately" challenging. Several flag officers, however, commented on the impact of world and domestic affairs on racial/ethnic tensions. Flag officers did not see unequal punishment of minority versus majority personnel for similar offenses as a significant problem. Only 16 percent saw this as a "great" or "extreme" challenge. Several flag officers commented that they were concerned about the Navy spouse, and treatment of single parents. Several also cited the importance of involving more Whites in the EO program and the necessity of staff training for EO personnel. Others identified bureaucratic delays in the civilian EEO program as well as the cost of settling EEO claims. They also mentioned various civilian age and handicap issues.

Degree Navy EO Program Effective/Understood

When asked questions on the degree to which the Navy's current EO program was understood or effective, respondents thought it was only moderately effective, with 51-55 percent giving the program "moderately effective" marks. The highest rating in this series was that the EO program was "moderately" effective in deterring racism/sexism; this item was rated as "highly" effective by only 39 percent of the flag officers. Only 38 percent of flag officers thought the Navy's EO program was "greatly" or "completely" understood, while 12 percent believed it was understood only to a small degree. Twenty-one percent of respondents felt human relations education attempts were ineffective in improving working relations. And 26 percent were displeased with the program's ability to provide effectively trained EO staff personnel. Thus, it appears that most flag officers believed the Navy's EO program had been only moderately effective and 10 to 26 percent were critical of the program.

Degree Flag Officers Prepared to Provide Equal Opportunity and Treatment

Seventy-five percent of flag officers see themselves as "greatly" or "completely" prepared to effectively provide equal opportunity to those under their leadership. Only three percent felt their military education, training, and experience did not adequately prepare them to provide equal opportunity to their subordinates. Thus, 97 percent felt at least moderately prepared to provide equal opportunity to those under their leadership.

Degree Proposed Seminar Objectives on Target

Three-fourths or more of flag officers agreed with all proposed EO Management Seminar objectives. Many seconded and amplified the basic objectives proposed through narrative comments. The percent which agreed that seminar objectives should include the following is listed below:

- Understand what types of flag-level EO support can create a more productive work climate (94% agree)
- Identify current EO issues in the Navy (92% agree)
- Understand the difference between command versus sailor level perceptions of equal opportunity climate, and the reason for these differences (91% agree)
- Understand the purpose and requirements of the Human Goals Statement, Navy EO orders, and Navy's new Command Managed Equal Opportunity Program (84% agree)
- Identify perceptual and statistical indicators which diagnose the equal opportunity health of a command--methods of determining the EO climate (80% agree)
- Understand the difference between a reactive versus a proactive equal opportunity program (78% agree)
- Understand the consequences of incorrectly reading the equal opportunity climate in the military in recent history (75% agree)

Thus, it appears that all objectives are supported by 75 percent or more of flag officers. The seminar appears to be on target.

Additional Seminar Topics

Several write-in comments indicated that admirals were also interested in demographic projections, language issues, utilization of women, fraternization policy (senior/junior; male/female), proven EO mangement techniques, and the qualifications and expectations commanding officers should have of EO personnel. Several also commented on course methodology, asking that concrete examples in concise English be used. Others asked for frank, honest assessment by knowledgeable members of various minority communities.

Flag Officer General Comments

Under general narrative comments, the following issues were addressed:

- The importance of the commanding officer assuming responsibility for EO
- Use of the chain of command to solve EO issues, and the fact that leadership must solve the problems
- Comments about CMEO--its lack of clarity, but moving in right direction
- Problems/strengths of the EO program and its implementation
- Continued discrimination, particularly involving women
- Need for a standardized and honest method for determining the EO climate
- On the proposed EO Management Seminar

Flag officers suggested case studies to learn from mistakes of others and asked that the seminar not be too basic but rather be "short and poignant." A few (9) thought that the seminar was not necessary.

Demographic Differences in Survey Responses

Although flag officers differed in their opinions, there appeared to be no large systematic differences between major types of assignments or commands. Only small differences between majority flag officers and minority flag officers as a group were identified. Because the minority flag officer population is so small compared with the White male majority, it was difficult to determine if each minority flag officer group had a different opinion in answering the survey because of their minority status or whether the two or three members of each minority group differed on individual questions from the majority as much as any two or three White male flag officers would.

Another question which could be asked of these data is: Did respondents who felt the EO program had served its usefulness and should be eliminated respond differently from those who believed it should be retained? Flag officers who wanted to eliminate the EO program differed significantly from others in their responses throughout the survey. Those that wanted the EO program done away with minimized the intergroup relations needing attention in the Navy, when compared with other respondents. They also felt EO was less important to accomplishing the Navy mission. Except for the commanding officer's role in modeling active support of EO and resolving EO issues in the chain of command, they felt all other elements of an EO program were less important. Regarding Navy EO challenges of the 1980s and 1990s, they minimized problems in the areas of accessions, upward mobility, and off-base treatment. Flag officers who wanted to eliminate the EO program, perceived that the Navy deals with all forms of discrimination more than those who want to maintain an EO program. They also thought the Navy assures equity more than those who see a need for an EO program. And finally, a larger percentage of those who want to eliminate the EO program disagreed with the proposed objectives for the Flag Officer EO Management Seminar. Thus, it appears that the answer a person gave to this question unlocks a cluster of related perceptions. The degree one perceives a need for the Navy EO program may be associated with the degree he/she believes EO issues appear to be concerns and the degree he/she believes the Navy already assures equity.

DISCUSSION

One of the strongest impressions conveyed from survey answers and narrative comments is the need to place equal opportunity within the chain of command and for commanding officers to take responsibility for affording equal opportunity to their subordinates. Flag officers believed that EO, interpersonal relationships, and mission readiness are inherently intertwined. Although 15-26 percent do not support the need for an EO program, 74 percent do see the need, and substantiated their opinion with additional written comments. Flag officers believed that while much has improved about the equal opportunity climate, the EO program has only been moderately effective; they appeared ready for a change in the program. Admirals were concerned about EO issues and what future demographic changes will bring to the Navy. They were particularly concerned about the utilization of women in nontraditional roles.

Flag officers believed the Navy is assuring equity to a far greater degree than do minority and female junior enlisted personnel. When one contrasts the high degree to which flag officers believed the Navy assures equity in promotions and military justice, and flag officers perceived their ability to provide equal opportunity to subordinates, to the opinions of minority and female personnel (as gathered by Human Resource Mangement surveys), one is struck by the difference. Minority and female sailors, particularly in the lower grades, were dissatisfied with equal opportunity and equity issues. Career minority and female sailors eventually change their attitudes toward equal opportunity and treatment, but the disparity exists with both their White male peers and with flag officers, at least through the grades of E-5 or E-6. Differences in the severity of judicial punishment, highlighted by the NPRDC report (Conway, 1983) on courts-martial rates and punishment, also contrast with flag officers opinions.

Seventy-five to 94 percent of flag officers believed the objectives devised by DEOMI are suitable for the Flag Officer EO Management Seminar. Although a few wrote reasons why they felt such training was unnecessary, it appears that 75 to 94 percent would support training on the proposed objectives. Since CMEO is newly implemented, it appears the near future is an opportune time for such training to clarify the program and sensitize flag officers to today's equal opportunity issues.

REFERENCES

1. Crawford, K. S., & Thomas, E. D., Human Resource Management and Non-judicial Punishment Rates on Navy Ships (NPRDC TR 76-5). San Diego: Navy Personnel Research and Development Center, August 1975. (AD-A013 336)

2. Conway, Susan W., Effects of Race and Gender on Court-Martial Rates and Punishment (NPRDC Special Report 83-20). San Diego: Navy Personnel Research and Development Center, March 1983.

3. Drexler, J. A., The Human Resource Management Survey: An Item Analysis: (Tech Rep.). Ann Arbor: University of Michigan, Institute for Social Research and Development Center, August 1975. (AD-A013 336)

4. Durning, K. P., How Enlisted Women and Men View the Navy Organization (NPRDC Tech. Rep. 82-23). San Diego: Navy Personnel Research and Development Center, January 1982. (AD-A110 116)

5. Thomas, Patricia J., Conway, Susan W., Racial/Ethnic and Gender Differences in Responses to the Human Resource Management Survey of Personnel Assigned to the Atlantic and Pacific Fleets (NPRDC SR 83-10). San Diego: Navy Personnel Research and Development Center, January 1983.

THE RELATIONSHIP AMONG VALUES, SITUATIONAL LEADERSHIP, AND OCCUPATIONAL SPECIALTY: A STUDY OF USAF FIELD GRADE OFFICERS

Lynville E. Taber
Air University

Abstract

The significance of values as a determinant of behavior has been advocated by behavioral science researchers and many authors agree that values impact decision making. A parallel issue is the notion that values play a major role in occupational choice. The subjects in this study are 199 USAF field grade officers. Subjects completed The Study of Values, the Leadership Effectiveness and Adaptability Description, and reported their Air Force Specialty Code. A canonical correlation analysis procedure revealed that among the variables of values, situational leadership styles, and occupational specialty, there were no statistically significant ($p<.05$) relationships.

Introduction

Background

The significance of values as a determinant of behavior has been espoused by numerous writers and behavioral science researchers for several years (Allport, 1935; Bandura, 1977). Wrenn (1980), a prolific writer in the human development field, asserts that, "The study and understanding of how values influence human behavior is of great importance." (p. 583) Many authors agree that values are the pivotal axis around which decisions are made and allude to the tremendous influence that values have in terms of an individuals's leadership behavior (Fiedler, 1967; Herzberg, 1979).

Conversely, some writers, (e.g., Adler, 1956; Hersey and Blanchard, 1982) take a different position. These authors argue that one can much more easily predict an individual's values/attitudes by observing their behavior. But, because of the extremely complex nature of both leadership behavior and individual's values, these two phenomena are particularly difficult to study/measure, especially conjointly. The difficulty in measurement may account for the disagreement in the literature regarding the power of values in terms of leadership attributes or leadership behavior (Bixenstine, 1976; Robinson & Shaver, 1973).

Another issue is the probability that values play a major role in choosing an occupation. Numerous researchers have concluded that values influence the selection of one's vocation and are instrumental in the success within that vocation (e.g., Massey, 1980; Power, 1980).

The Central Issue

In several books, reports, and journal articles written recently, the subject of leadership is seemingly connected with values or value systems in one context or another. The leadership-values connection appears to be endemic. That is, it seems impossible for American authors to conceptualize the activity of leadership without including a discussion of the value contstructs operating within the leader (Bennis, 1984; Huntington, 1957; Peters & Waterman, 1982; Van Fleet, 1984; Zaleznik, 1984).

To explore the issues raised in the above paragraphs, this research examined the relationship among values, situational leadership, and occupational specialty in a sample of U.S. Air Force field grade officers.

Method

Sample

The subjects for this study were 199 USAF officers (majors) attending Air Command and Staff College (ACSC), Air University, Maxwell AFB, AL. Officers were randomly selected from a population of 385 USAF majors attending ACSC. All subjects of this study were male, and averaged 36 years of age. The subjects' occupational specialties (Air Force Specialty Code-AFSC) were reported, and appeared to approximate all AFSCs for the USAF officer corps. The sample included 103 (51.8%) "rated" (pilot or navigator) subjects; 28 subjects (14.1%) within scientific and technological occupational specialties; and, 68 (34.2%) of the subjects were in a category classified as "non-rated and non-scientific."

Measures

Value data were collected using the *Study of Values* (SOV) (Allport, Vernon, & Lindzey, 1960). The SOV aims to measure the relative prominence of six basic interests in an individual's personality. The classification of these six value sets is based on Spranger's *Types of Men* (1928) which describe the following dominant interests: *Theoretical* - emperical, critical, and intellectual; *Economic* - useful and practical; *Aesthetic* - form, grace, and symmetry; *Social* - love of people, kind, unselfish, and sympathetic; *Political* - power, competition, and struggle; and *Religious* - love of unity and affirms life.

Each subject also completed the *Leadership Effectiveness and Adaptability Description* (LEAD) (Hersey & Blanchard, 1974), providing data regarding measures of leadership styles in differing situations. The LEAD purports to measures four leadership styles (telling, selling, participating, and delegating) and the adaptability or effectiveness of the leader, i.e., the ability to change leadership styles depending upon the situation and the maturity level of the followers.

Procedures

The SOV and LEAD instruments were assigned as homework to increase student learning during the leadership block of instruction. The SOV was assigned first and the LEAD assignment trailed in the curriculum by two days. The SOV was accompanied by a demographic sheet enabling the collection of limited demographic data.

Data Analysis

Because of the number of variables involved in the study--values = 6, situational leadership = 4, and three AFSC groupings--a canonical correlation analysis statistical procedure was employed. Canonical correlation analysis is the most generalized member of the family of multivariate statistical techniques and has as a goal the determination of the primary independent dimensions which relate one set of variables to another set of variables (Bartlett, 1938).

Results

For the SOV, the three highest means were on the Political, Economic, and Theoretical scales; in that order. For the LEAD, the subjects reported a predominant style of "selling" (high task and high relationship behavior), with "participating" (high relationship and low task) being the alternate style of leadership. The least frequent style of leadership reported was "telling" (high task, low relationship).

Two canonical correlation analyses were attempted. The first related the six SOV scores with the four leadership style scores and AFSC. The second related the six SOV scores with the LEAD's four adaptability scores and AFSC. On the two relate lists, no canonical correlations between or among the variables were found at the $p<.05$ level of significance. (See Table 1, next page, for Inter-Item Correlation Coefficients. Other tables and expanded reference lists are available from the author upon request.)

Discussion

The lack of statistical significance when correlating values, situational leadership, and AFSC variables is puzzling in light of the literature reporting the strong influence of values upon human behavior. It is possible that the two instruments employed in the study were valid; that is, exact in measuring only what they were designed to measure, and that a correlational finding between them is not feasible. Perhaps, also, the subjects in this study (the top 25-30% of majors in the USAF) are so well educated and trained in the art of leadership that they do not allow their personal value systems to interfere with the decision making processes. A third notion is that, regardless of the AFSC, the subjects in this study were able to transcend their occupational specialty biases and choose a viable alternative when responding to the LEAD's situational descriptions.

Conclusions

The findings in this study suggest that neither values nor occupational specialty play a major role in situational leadership. Caution is urged, however, in generalizing these findings to USAF field grade officers on the whole. Further research is required, perhaps utilizing different value/leadership instruments. A comparative study, involving executive-level civilians in similar occupational specialties is also suggested.

Table 1

SOV & LEAD Inter-Item Correlation Coefficients

	Theor.	Econ.	Aesth.	Social.	Polit.
Theoretical	1.000	0.070	0.013	-0.281	-0.067
Economic		1.000	-0.278	-0.267	0.208
Aesthetic			1.000	-0.054	-0.269
Social				1.000	-0.353
Political					1.000
Religious	-0.461	-0.456	-0.270	0.028	-0.285
Telling	0.088	-0.023	-0.002	-0.129	-0.088
Selling	-0.115	0.040	-0.015	0.020	0.015
Participating	-0.101	-0.012	0.035	0.125	-0.102
Delegating	0.283	-0.025	-0.013	-0.103	0.038
Effectiveness	0.056	0.012	-0.040	0.061	0.002

	Relig.	Tell	Sell	Partic.	Deleg.
Theoretical	-0.461	0.088	-0.115	-0.101	0.283
Economic	-0.456	-0.023	0.040	-0.012	-0.025
Aesthetic	-0.270	-0.002	-0.015	0.035	-0.013
Social	0.028	-0.129	0.020	0.125	-0.103
Political	-0.285	0.087	0.015	-0.102	0.038
Religious	1.000	-0.016	0.031	0.029	-0.097
Telling		1.000	-0.414	-0.301	-0.017
Selling			1.000	-0.580	-0.366
Participating				1.000	-0.153
Delegating					1.000
Effectiveness	-0.015	-0.118	-0.317	0.216	0.392

	Effect.
Theoretical	0.056
Economic	0.012
Aesthetic	-0.040
Social	0.061
Political	0.002
Religious	-0.015
Telling	-0.118
Selling	-0.317
Participating	0.216
Delegating	0.392
Effectiveness	1.000

References

Adler, F. (1956). The value concept in sociology. American Journal of Sociology, 62, 272-279.

Allport, G. W. (1935). Attitudes. In C. Murchinson (Ed.), Handbook of social psychology. Worchester, MA: Clark University Press.

Allport, G. W., Vernon, P. E., & Lindzey, G. (1960). Study of Values. (3d ed.). Boston: Boston University.

Bandura, A. (1977). Social learning theory. Englewood Cliffs, NY: Prentice-Hall.

Bartlett, M. S. (1938). Further aspects of the theory of multiple regression. Proceedings of the Cambridge Philosophical Society, 34, 33-40.

Bennis, W. G. (1984). The 4 competencies of leadership. Training and Development Journal, 38, 15-19.

Bixenstine, E. (1976). The value-fact antithesis in behavioral science. Journal of Humanistic Psychology, 16, 35-67.

Fiedler, F. E. (1967). A theory of leadership effectiveness. New York: McGraw-Hill.

Hersey, P., & Blanchard, K. H. (1974). So you want to know your leadership style? Training and Development Journal. (February).

Hersey, P., & Blanchard, K. H. (1982). Management of organizational behavior: Utilizing human resources (4th ed.). New Jersey: Prentice-Hall.

Herzberg, F. H. (1979). Herzberg on motivation for the '80s: Piecing together generations of values. Industry Week. (October).

Huntington, S. P. (1957). The soldier and the state: The theory and politics of civil-military relations. Cambridge, MA: Harvard University Press.

Massey, M. E. (1980). The people puzzle - What you are is. Reston, VA: Reston.

Peters, T. J. & Waterman, J. H., Jr. (1982). In search of excellence: Lessons from America's best-run companies. New York: Harper & Row.

Power, C. (1980). Further reflections on values assessments. Personnel and Guidance Journal, 58, 616-617.

Robinson, J. P., & Shaver, P. R. (1973). Measures of social psychological attitudes. Ann Arbor, MI: University of Michigan Institute for Social Research.

Spranger, E. (1928). Types of men (P. J. Pigors, Ed. and Trans.). New York: Stechert-Hafner.

Van Fleet, D. D. (1984). Organizational differences in critical leader behavior: Industry and military. In R. L. Taylor & W. E. Rosenbach (Eds.), Military leadership: In pursuit of excellence (pp. 120-127). Boulder, CO: Westview.

Wrenn, C. G. (1980). A commentary on the role of values in the counseling situation. The Personnel and Guidance Journal, 58, 582-583.

Zaleznick, A. (1984). The leadership gap. In R. L. Taylor & W. E. Rosenbach (EDs.), Military leadership: In pursuit of excellence (pp. 86-96). Boulder, CO: Westview.

A Longitudinal Study of the Job Attitudes of U.S. Air Force Pilots: Revisited

Robert A. Gregory
Center for Creative Leadership and
United States Air Force Academy

Abstract

The relationship of job attitudes to retention of Air Force (AF) pilots has been an area of study for several years. Three administrations of the Job Diagnostic Survey (Hackman & Oldham, 1975) have been accomplished, and the data have supported changes within the AF as well as documenting the impact of external factors on pilot retention. Job characteristics, pay and benefits and external factors such as the economy and airline hiring practices have had major impacts on pilot retention.

Introduction

Rosenbach (1982) reported on a replication of a study (Gregory & Rosenbach, 1981) in which the job attitudes of U.S. Air Force (USAF) pilots were examined with the hope of finding significant areas that might be related to retention in the AF. The initial research, which compared USAF pilots to pilots in a major U.S. airline, was conducted in 1979; the follow-up occurred in 1982. The current report includes a third iteration of that effort which was completed in 1984. While numerous changes to policies and procedures within the USAF had occurred between 1979 and 1982, there had been a stablization between 1982 and 1984. For example, no dramatic personnel policies, pay raises, bonuses, etc. had been implemented. So, too, had the external environment stablized in those years; the airline hiring had been relatively low and constant (furloughed pilots were being recalled, reducing that pool) and the economy had been on a fairly stable, predictable course.

There were indications, however, in 1984, that some areas which should be of concern were re-emerging. Retention was beginning to show a decrease; the number of pilots establishing dates-of-separation was on the rise. In addition, the economy was beginning to show a healthy improvement from the economic doldrums of the past year or two. Perhaps more importantly, the benign neglect being experienced by USAF pilots seemed to be showing little promise of improving. The retirement system was being seriously threatened; the pay raises and delays in those raises had decreased and increased, respectively; and the swing of the pendelum in society to a more conservative stance seemed to be eagerly embraced by public sector leadership--seemingly resulting in less concern about benefits for the public servant. Not to be

neglected, were convincing reports of dramatic increases in airline hirings. To examine the dynamic relationship between job attitudes and some of these changes it seemed essential to take a "snapshot" in 1984. If, indeed, there was a significant change in various attitudes, policy-makers might be able to avoid the kind of crisis that was experienced with the extremely low retention rate of the late 1970's.

In 1979 commercial airline pilots had significantly more intrinsically motivating jobs, and they were much more satisfied with their general job context than USAF pilots (Gregory & Rosenbach, 1981). Much improvement in USAF jobs and context variables had occurred by 1982 (Rosenbach, 1982). The economy's downturn (high unemployment and high interest rates) coupled with commercial airline pilot furloughs were thought to have had major impacts on retention of USAF pilots (Ginovsky, 1982a). Even though significant favorable (1982, 1982a) changes in job attitudes had occurred from 1979 to 1982, a great deal of dissatisfaction continued to be reported in two major areas: personnel policies; and leadership, management and supervision. In 1982, Rosenbach concluded that the improved outlook on retention was a result of more motivating jobs and improved benefits coupled with a depressed economy. Retention, it was predicted, should remain higher than 1979 levels, but continuing efforts to improve the motivating potential of USAF pilots' jobs and maintaining equitable pay and benefits are crucial to avoid a recurrance of the 1979 crisis. The current study is a part of the ongoing effort to examine that conclusion and better understand the dynamics of USAF pilot job attitudes.

Method

Subjects

The subjects were a stratified random sample of USAF pilots (first lieutenants, captains, majors, and lieutenant colonels) from the four major flying commands of the AF (Air Training Command (ATC) Military Airlift Command (MAC), Strategic Air Command (SAC) and Tactical Air Command (TAC), in the continental U.S.; two overseas commands were added in the 1984 study. The 1984 sample size was 2405. The 1982 and 1979 samples were 1582 and 3167, respectively.

Procedure

The primary data gathering mechanism was a paper-and-pencil survey which was mailed to each pilot along with a machine-readable response sheet and a return envelope. Supplemental information was obtained from letters and comments submitted by the subjects and telephone interviews.

The survey instrument was a modified version of the Job Diagnostic Survey (JDS) (Hackman & Oldham, 1975). Attached to the survey was a comment sheet with a note encouraging the pilots to write comments on any aspect of their job or organization they wished. The JDS provided measures of the five core job characteristics as follows: skill variety, task identity, task significance, antonomy and feedback from the job.

The scores for the five job characteristics (which may range from 1 to 7) can be combined into a single index, the Motivating Potential Score (MPS) which reflects the overall potential of a job to foster internal work motivation on the part of job incumbents. The resultant MPS score may vary from a low of 1 to a high of 343. The national average for all jobs is 128 (Hackman & Oldham, 1980). The JDS also measures the job outcome variables: general job satisfaction and satisfaction with growth. Additionally, the instrument measured the following moderator variables: growth need strength, satisfaction with pay, satisfaction with security and satisfaction with supervision. Areas added to the basic JDS included factors having a career influence on members, spouse satisfaction and career intent.

Telephone interviews were conducted with those pilots who identified themselves and indicated a willingness to elaborate on their written comments. Content analyses were performed on both the interview data as well as data from comment sheets and letters.

RESULTS

Of the 4249 pilots who were sent the survey, 2405 completed and returned it for a response rate of 56.6 percent.

Table 1: Variables, Means and Standard Deviations for Job Attitudes of USAF Pilots

Variable	1979 n=3167		1982 1582		1984 2405	
	M	SD	M	SD	M	SD
Skill Variety	5.3	1.3	5.4	1.2	5.5	1.2
Task Identity	4.7	1.4	4.9	1.3	4.9	1.4
Task Significance	5.5	1.3	5.7	1.2	5.4	1.2
Autonomy	4.4	1.5	4.6	1.4	4.9	1.4
Feedback from Job	4.8	1.2	5.1	1.2	5.0	1.2
Motivating Potential Score (MPS)	121	71.2	134	70.1	139	72.7
General Job Satisfaction	4.9	1.5	5.3	1.4	5.2	1.4
Satisfaction with Growth	4.7	1.4	5.1	1.2	5.1	1.3
Satisfaction with Pay	3.0	1.6	4.5	1.6	4.4	1.6
Satisfaction with Security	4.3	1.7	4.9	1.4	4.7	1.5
Feedback from Supervisors	4.2	1.4	4.4	1.5	4.3	1.5
Satisfaction with Supervisors	4.9	1.5	5.0	1.4	5.0	1.4
Social Satisfaction	5.5	0.9	5.6	0.8	5.5	0.9
Growth Need Strength	6.1	0.9	6.1	0.9	6.1	0.9

Table 1 shows the results for the survey on the JDS measures for the three years. The only factors on which the 1984 group increased were skill variety and autonomy; these increases more than counterbalanced the decreases in feedback from job and tasks significance, resulting in a slight increase in the MPS. Other decreases were observed in general job satisfaction, satisfaction with security, feedback from supervisors and social satisfaction. It is important to note that addition of the overseas commands inflated the 1984 numbers slightly; they were above the overall mean on practically all measures, thereby causing the 1984 picture to appear slightly more favorable.

Content Analysis of Comments

Almost half the respondents also submitted written comments. About 20 percent of the comments in 1984 were favorable (an increase from prior years). Table 2 shows the results of the content analysis of those comments.

Table 2: Comments from USAF Pilots Concerning Their Jobs

Favorable 1979	1982	1984
1. The job itself - flying	1. The job itself - flying	1. The job itself
2. The survey - interest of leadership	2. Air Force way of life	2. Pay and benefits
3. Air Force people - co-worker	3. Air Force people - co-worker	3. Flying
		4. Personnel
		5. Patriotism/opportunity to serve country
		6. People/colleagues
		7. Travel

Unfavorable 1979	1982	1984
1. Personnel policies	1. Personnel policies	1. Personnel (promotions/assignments)
2. Leadership/management/ supervision	2. Leadership/management/ supervision	2. Pay and benefits
3. Pay and benefits	3. Family considerations	3. Leadership/management/supervision
4. Long hours	4. Shortage of resources	4. PCS
5. Additional duties		5. Job dissatisfaction
6. Shortage of resources		6. Long hours
		7. Additional duties

The job and flying came out consistently as highly favorable aspects of AF life as did AF people/colleagues. Pay and benefits were often mentioned in 1984 as "not too bad" which was categorized as a favorable response. Re-emerging very strongly in 1984 as an unfavorable factor, however, was pay and benefits; this comment generally related to concern about low raises, the cost of living and, more strongly, the perceived threats to the current retirement system. Personnel policies and leadership, management and supervision remained among the most frequent unfavorable areas mentioned.

Additional Data

Of a list of 15 factors generally perceived as favorable toward making the AF a career, the top five (in descending order of importance) were: job challenges (24.9 percent), retirement (18.4 percent), assignment choice (10.7 percent), promotions (9.8 percent), and pay (8.7 percent). Of the same 15 factors those which "if changed from their current status would make me consider leaving the AF" included: retirement (55.1 percent), pay (11.1 percent), job challenge (8.6 percent), promotions (5.5 percent), and fringe benefits (3.1 percent).

Though it varied considerably among commands there was an increase in the stated propensity to leave the AF by the 1984 respondents compared to the 1982 respondents. In 1984 there was almost a three percent increase in the number of captains who reported that they were either "definitely planning" or "thinking about" leaving the AF (with much more dramatic increases in TAC and SAC). The percent of lieutenant colonels expressing an intent to leave at 20 years increased dramatically in MAC and somewhat in ATC while it dropped slightly in SAC and TAC; the overall percentage intending to separate earlier went up approximately one and one half percent.

Discussion

The results of the JDS measure of job characterics show considerable stability from 1982 to 1984; most of the dramatic improvements from 1979 and 1982 have been maintained. While the significant drop in task significance deserves further study the increase in autonomy is promising and suggests that pilots are being allowed more decision-making in their jobs. The fact that general job satisfaction and satisfaction with pay and security have dropped are causes for serious concern. This indicates that pilots have a global

feeling that their jobs are not as satisfying as they were in 1982 and that their future is not as secure as they'd like it to be. This insecurity undoubtedly ties in directly with the written concerns about pay and benefits (the "things are o.k. now, but looking to the future..." attitude) and the threats to the retirement system indicated in both the quantitative data and the written comments.

Perhaps the most serious cause for concern centers around the significant increase in the percentage of captains who indicate that they are planning to leave the AF. It seems that the concerns about the temporary nature of higher retention over the past few years were well founded. With the recent signs of an improving economy, the furloughed pilots being rehired, and predictions that the airlines were going on a hiring spree coupled with a stagnation in improvements in the life of AF pilots along with very real threats to future security, decreases in retention seem inevitable.

As a result of the 1979 retention crisis and other factors, numerous changes were instituted in the AF to improve the life of AF pilots. More flying hours were available; some commands made significant efforts in improving autonomy for pilots; and pay and benefits improved. Many of the changes were initiated by and within the AF. Retention improved. It is quite clear that a complex combination of factors impact AF pilot retention; some of these are external to and uncontrolled by the Air Force, but a number of factors are <u>not</u> controlled by Congress and economic conditions. Constant vigilance to all factors and appropriate actions to correct problems over which the AF can exercise control is critical. Anytime an organization depends on external factors to take care of its problems it must be ready for a varied, unpredictable impact. The AF is no exception.

REFERENCES

Ginovsky, J. (1982a, July 26). Hiring of pilots drops as airline industry slides. <u>Air Force Times</u>, p. 7.

Ginovsky, J. (1982, June 28). Fewer military pilots seeking airline jobs. <u>Air Force Times</u>, p. 2.

Gregory, R. A., & Rosenbach, W. E. (1981, August). <u>Retention of USAF Officers</u>. Paper presented at the Eighty-Ninth Annual Convention of the American Psychological Association, Los Angeles, CA.

Hackman, J. R., & Oldham, G. R. (1980). <u>Work redesign</u>. Reading: Addison-Wesley.

Hackman, J. R., & Oldham, G. R. (1975). Development of the Job Diagnostic Survey. <u>Journal of Applied Psychology</u>, <u>60</u>, 159-170.

Rosenbach, W. E. (1982, August). <u>A longitudinal study of job attitudes of U.S. Air Force pilots</u>. Paper presented at the ninetieth annual convention of the American Psychological Association, Washington, D.C.

Rosenbach, W. E., & Gregory, R. A. (1980). Job attitudes of commerical and USAF pilots. <u>Armed Forces and Society</u>, <u>8</u>, 615-628.

PRODUCTIVITY IN THE USAF SYSTEMS COMMAND

Vince Luchsinger, Maj.Gen. USAFR
Mob. Asst. to Commander, USAF Systems Command
& University of Baltimore

INTRODUCTION TO THE COMMAND

The primary mission of the Air Force Systems Command (AFSC) is to advance aerospace technology, apply it to operational aerospace systems development and improvement, and acquire qualitatively superior, cost-effective, and logistically supportable aerospace systems. The command has been involved in space missions from the earliest United States ventures, going back to the early 1950's in exploratory work by General Bernard Schriever a past Systems Command commander.

The Air Force Systems Command also supports the major space responsibilities of the Department of Defense, including basic and applied research, development, test and engineering of satellites, space launches and missions, boosters, space probes, and related space systems. The Command is presently heavily involved in the research activities of the Space Defense Initiatives program, under the command of a space veteran of AFSC experience, General Abrahamson. In addition, AFSC supports many NASA programs and projects that operate under agreements between the Department of Defense and NASA.

AFSC Resources

The Command is one of the larger Air Force Commands. In scope of financial activity, it would rank sixth among the Fortune 500 of United States corporations. The personnel ranks of AFSC list 27,500 military and 29,500 civilian personnel. The nature of its research, development, test, and acquisition means that AFSC is the primary Air Force employer of scientists, engineers, and technically oriented personnel.

In the current 1985 fiscal year, Systems Command is managing approximately $38 billion. Within that amount, $23.1 billion goes for procurement of aircraft ($16 billion), missiles ($5.3 billion), and other equipment ($1.8 billion). Beyond that, $10 billion is applied to research, development, test, and evaluation (RDT&E), $1.5 billion for operation and maintenance, and $540 for military construction. The remaining $3 billion comprises foreign military sales ($1 billion), reimbursables ($1.2 billion), and military pay ($900 million).

The magnitude of the resources entrusted to AFSC dictates a heavy responsibility for productivity in discharging its missions to ensure the best output for the resources involved. As a vital

segment of the Air Force structure, AFSC administers thirty-eight percent of the total Air Force budget, while utilizing only 6.5 percent of the personnel of the Air Force. As a final indicator of the key role of the Systems Command, AFSC currently administers 29,000 active contracts valued at approximately $108 billion.

PEOPLE PRODUCTIVITY INITIATIVES

One vital segment of the productivity program of AFSC rests in people programs. In August 1984, General Lawrence Skantze assumed command of AFSC and indicated a concern for productivity initiatives within the Command. While entrusted with thousands of talented people and a large fiscal and acquisition responsibility by the Air Force, General Skantze saw that executing the mission of the Command would soon involve "doing more with less". The vast array of program responsibilities with increasing costs would test the ability of the Command to respond.

Changing the Culture

Contemporary management research and writing attests to the need for assessing the culture of organizations to verify goals and values in pursuing goals of excellence, change, or renewal. The culture change process is at work in AFSC. As in often the case in a change of command, the new commander will establish his style and working expectations in carrying the organization to the accomplishment of missions. Since AFSC is a linking pin between the using operational commands of the Air Force and the community of research and development (R&D), acquisition, and civilian contracting firms, the Command finds itself a military organization with multiple ties to segments of the civilian world.

General Skantze [2] has placed emphasis on managing the Air Force acquisition role in a "business-like" manner. This means that AFSC military and civilian personnel not only must meet military standards, but must be able to perform creditably in dealings and negotiations with civilian contracting organizations and R&D enterprises. AFSC personnel performing well will best support their colleagues in the operational and combat commands. "Buying smartly" would become more than a catchy slogan.

Managing intelligently includes a deep concern for quality. Quality assurance is a major element of AFSC acquisition and contract management programs. Accompanying quality in staff work and business management programs of the command are endorsed as a corollary of the desired quality sought in procurement activities.

People Programs

An array of ventures to involve Command personnel in improving productivity and quality operations includes:

1. Suggestion Programs which have been emphasized to generate ideas submitted to proposal improvements. The quantity and quality of suggestions has been gratifying in response to Command emphasis. While a common management tool, suggestion systems have reflected commitment of organization members.

2. Sensing sessions with senior officers from field commands to assess problems and opportunities in those organizations, as well as to examine inter-organization relationships. Input from those key personnel has been instrumental is working out kinks in operations, and in reducing we-they confrontations. This was one of many innovations with behavioral science flavor.

3. Off-site sessions by general officers and key staff to review objectives, values, weaknesses, strengths, and opportunities which face the Command. These work meetings (in civilian clothes) promote creative and cohesive options for future Command operations.

4. An organizational survey of the AFSC headquarters was executed by the Air Force Leadership and Management Development Center to provide data to managers of organizational components. These data cover job satisfaction, attitudes toward work group and supervisors, organizational communications, and other topics. Those findings should provide opportunities for improving management and productivity in the headquarters.

5. A civilian personnel work force effectiveness study group has been convened with representation from all field units. This group is generating initiatives for optimizing the contributions of the civilian work force, an important segment of the AFSC family.

6. Enhanced career development programs are being put in place to provide career pathing for personnel. The career satisfaction and progress is important for the support of individual and work group productivity, as well as to provide work force stability.

7. Decentralization, whenever possible, as a means of establishing the importance of the "work place" and the "work unit". Senior officers, such as General Larry Welch, a former vice chief of staff [4], have endorsed decentralization as a move to promote participation of personnel at all levels in pursuing quality and productivity objectives.

The foregoing are just examples of an emphasis on people programs in AFSC which are designed to promote an awareness of quality and productivity in the culture of the Command. The entire Command has been sensitized to the need for greater

attention to the use of people and oth... ...es to meet goals of each organization in AFSC. Working "s... ter" and "better" is encouraged as a means of dealing with the continually increasing workload of the Command.

RESOURCE PRODUCTIVITY

Since AFSC is entrusted with a major segment of the Air Force buying role, the use of sophisticated and contemporary tools and techniques are required to manage the development, test, and acquisition activities of the Command. Some of the major productivity impacts follow.

Management Information Systems

To administer the functions inherent in the AFSC mission, the management of information is essential. Vast amounts of data must be selectively acquired, stored, analyzed, and presented to decision makers to manage the Command. Executive Information System (EIS) capability is availabe within and between elements of the command to maintain real-time communications. While electronic mail and filing (among other electronic capabilities) are common place in industry, the military necessities of security, readiness, and global coverage accentuate the EIS mission.

A new staff directorate (SI) was formed in 1984 to deal with computer resources and management information systems. The system architecture is being developed, with growth and versatility of function a key element. This capability includes use and networks of micro-computers, as well as networking with main-frame and support computing facilities. It is worth noting that AFSC facilities include equipment ranging from micro-computers to super-computers. Needless to say, extensive design and training activity is involved to achieve and maintain the productivity that can be attained from electronic capabilities.

Planning

Better management for productivity involves improved planning with accompanying control systems. In the quest for more productivity in its mission areas, AFSC uses planning extensively. A prime example is the Space Technology Plan [3], a product of the Space Technology Center. This type planning has a strong impact on the Command and relationships with DOD, NASA, other government agencies and industry.

Space technology planning focuses all Air Force space technology investment and execution in support of future space mission requirements. Selected space technology development projects will pursue objectives in : on-board processing and hardened electronics, spacecraft autonomy, space power, and advanced military spacecraft capability.

Artificial intelligence, power cell research, and advanced space computer technology are hoped to provide increased spacecraft lifetime, memory, survivability and autonomy. The productivity of this research is tied to unique features of space needs. Some of those factors include:

1. Space systems have long lead times and long life times.

2. Technology development must start very early before systems are well defined.

3. There is a premium on accurate forecasting of technology needs.

The complexities of technologies of development, test, and production make the tasks of AFSC challenging. Partnerships are being forged with other agencies and industry to provide proactive approaches to managing for quality outcomes in a cost-effective manner.

CONCLUSION

A recent NASA report on Improving Quality and Productivity in Government and Industry [1] captures the essence of the AFSC quest for productivity in its activities. Quality of work, quality of work life, and quality of management are cornerstones for increasing productivity. Technology and production bases help provide the setting for quality of work in meeting standards and requirements for products and outputs.

People programs project commitment and work environment conditions which contribute to the success of the organization. Quality of management fosters leadership that has the technical and leadership skills to provide direction and feedback in guiding the organization to its objectives. The people of AFSC are working toward those quality conditions in the pursuit of productivity.

REFERENCES

[1] NASA, A Framework for Action: Improving Quality and Productivity in Government and Industry, December, 1984.

[2] Skantze, Lawrence. Personal communication. July, 1985.

[3] Space Technology Plan, Space Technology Center, Kirtland AFB, NM. 1985.

[4] Welch, Larry D. Management and Productivity. Remarks to the National Forum of Human Resource Planning, Baltimore, MD. May 9, 1985.

EFFECTS OF ABILITY AND MOTIVATION ON THE PERFORMANCE OF MECHANICS

BRUCE STERLING
UNITED STATES ARMY SOLDIER SUPPORT CENTER

ABSTRACT

Campbell (1976) postulates that an individual's performance is a function of both ability and motivation. Yet most studies focus on only one of these two areas. This study examined the relationship between ability, motivation and performance rankings of 277 junior enlisted mechanics. Results showed that job performance is positively related to both mechanics' general mechanical (GM) aptitude and to self ratings of motivation. These measures have a combined effect such as that of those with high aptitude and high motiation, less than 10% were ranked in the lowest performance category, while of those with high aptitude but low motivation, nearly 40% were thus ranked. This data suggests that having high aptitude soldiers alone does not ensure effective performance.

INTRODUCTION

Campbell (1976) states that an individual's performance is a function of both motivation and ability. There is a rich literature in industrial/ organizational psychology in each of these areas. However, most research on the role of motivation in human performance either ignores aptitude or measures its effects only to examine aptitude as an intervening variable. For example, Prichard, Hollenback and Deleo (1980), in examining the effects of contingent vs non-contingent reinforcement on performance, tested subjects on several aptitudes relevant to the task. However, they used this data merely to ensure no significant differences in aptitude existed between the groups, rather than to assess how aptitude and motivation influenced performance. Similarly, research concerning the effects of aptitude on performance do not consider the role of motivation (e.g., Ghiselli, 1973).

The roles of motivation and aptitude in human performance are of more than theoretical interest. The US Army has recently amassed evidence that general intelligence (as measured by a composite score derived from an intelligence test designed for the military) is positively associated with performance for a variety of jobs (military occupational specialties or MOS). For example, Scribner, Smith, Baldwin, and Phillips (1984) found that tanks crewed by higher mental category tank commanders or gunners fared better in tank gunnery exercises. The problem is that the number of soldiers in the higher mental categories is limited. Therefore it would be of some practical use to know what role motivation and aptitude play in determining performance; for instance, does a gunner who is high in motivation but low in aptitude perform similarly to a gunner low in motivation but high in aptitude? Can we "compensate" for low aptitude by motivation? Tziner and Eden (1985) recently conducted a study in the Isreali Defense Force examining the roles of crew member motivation and ability in tank crew performance. Main effects were

found for the motivation of two crew members and the ability of all three crew members studied, but no interactions were found between ability and motivation. However, data is not presented in such a manner that the above questions concerning "tradeoffs" between aptitude and motivation can be answered.

The purpose of this research is to examine the roles of both motivation and aptitude in determining soldier performance.

Method

Subjects. Subjects were 123 junior enlisted mechanics (MOS 63B, 63T, 63N, and 63H) in nine US Army battalions at two installations within the continental United States.

Instruments.

Aptitude: All soldiers, upon entering service, are given the Armed Services Vocational Aptitude Battery (ASVAB), a measure of several aptitudes. Results discussed here will be limited to the aptitude most relevant to this MOS, the general mechanical (GM) subscale. Soldiers' GM scores were obtained by entering soldiers' social security numbers (obtained from the units) into an Army data base.

Motivation: Eighty-one of these soldiers (all available when research was conducted) completed a 17 item survey, which contained a self-evaluation of job motivaton on a five point scale.

Performance: Measurement of performance was done using a procedure adapted from Stouffer, Lumsdaine and Lumsdaine (1949). In each company motor pool, the motor officer was asked by the interviewer (an Army officer) to indicate the mechanic he/she would be least willing to lose (solely in terms of job performance) if one had to be taken from the motor pool to help form a new unit. They were then asked to indicate the mechanic they would be most willing to lose. If there were 3 or 4 mechanics in the motor pool, the procedure ended there. If there were 5 or 6, two iterations of the above were conducted, and if 7 or 8, three iterations. In all cases, a top, bottom and middle group (roughly corresponding to top, middle and bottom thirds) was created.

Analyses: GM scores were divided into those scoring 100 (population average) or greater versus less than 100. Motivation scores were divided into those scoring at or above the median (top two categories) versus below the median (bottom three categories). Chi-square analyses were then performed for motivation-by-performance category, aptitude-by-performance category and aptitude-by-performance category for mechanics high versus low in motivation.

Results

Table 1 (below) shows that mechanics with GM scores of 100 or greater were more likely to be ranked in the top performance category, and less likely to be ranked in the bottom category, than mechanics with GM scores of less than 100.

Table 1

Mechanics' Ranking by General Mechanical (GM) Score

Ranking Category	GM Score: Less than 100	GM Score: Greater than or equal to 100
Top	N=11 25.6%	N=32 40%
Middle	N=9 20.9%	N=28 35%
Bottom	N=23 53.5%	N=20 25%

Chi-square = 9.99, Degrees of freedom = 2, P less than .01

Table 2 (below) shows that mechanics who rated their own job motivation in the bottom three categories were more than twice as likely to be ranked in the bottom performance category than those who rated their job motivation in the upper two categories.

Table 2

Mechanics' Rankings by Motivation

Ranking Category	Motivation: Somewhat/ Not very/Not at all Motivated	Motivation: Highly Motivated/ Motivated
Top	N=11 31.4%	N=16 34.8%
Middle	N=6 17%	N=20 43.5%
Bottom	N=18 51.5%	N=10 21.7%

Chi Square = 9.43, degrees of freedom = 2, P less than .01

Table 3 shows the combined effects of motivation and aptitude. Of those with high motivation and aptitude, less than 10% were ranked in the bottom performance category, while of those with high aptitude but low motivation, nearly four times that percentage were thus ranked. Further, of those high in motivation but low in aptitude less than half were ranked in the bottom performance category, while of those low in both factors, three-quarters were thus ranked.

Table 3

Mechanics' Ranking by Aptitude, as moderated by motivation

High Motivation

Ranking Category	GM Score: Less than 100	GM Score: Greater than or equal to 100
Top	N=2 15.4%	N=13 41.9%
Middle	N=5 38.5%	N=15 48.4%
Bottom	N=6 46.2%	N=3 9.7%

Chi-square = 8.05, Degrees of freedom = 2, P less than .05

Low Motivation

Ranking Category	GM Score: Less than 100	GM Score: Greater than or equal to 100
Top	N=1 8.3%	N=10 47.6%
Middle	N=2 16.7%	N=3 14.3%
Bottom	N=9 75%	N=8 38.1%

Chi-square = 5.58, Degrees of freedom = 2, P less than .07

Discussion

In answer to the specific question raised in the introduction, i.e., can one substitute a highly motivated low aptitude soldier for a less motivated high aptitude soldier, the answer appears to be no. About equal percentages of thse two groups are found in the bottom performance category, but three times the percentage of high aptitude persons are in the top performance category. However, this does not mean that motivation is unimportant. For both high and low aptitude soldiers, roughly an equal percentage of highly motivated and less motivated soldiers are found in the top performance category. However, a substantially larger percentage of less motivated soldiers, regardless of aptitude, are found in the bottom performance category. Thus the effects of high aptitude on performance are substantially reduced by low motivation.

Motivation is basically a situational variable. That is, a soldier may be highly motivated to perform in the field but not the classroom, or vice versa. One of the most important determinants of motivation is whether rewards are contingent on performance. For most Army performance measures, it is more likely that punishment is contingent on failure (e.g., bar to reenlistment for failing one's skill qualification test) than reward being contingent on success. Further, even this type of punishment does not affect the large population of first term soldiers who do not intend to reenlist. While the description of organizational behavior modification programs is beyond the scope of this paper (see Miller, 1978) programs linking rewards to performance (hence enhancing motivation) could be designed at unit and Army levels.

In summary, this research demonstrates that motivation cannot replace high aptitude as a determinant of performance, but motivation significantly enhances the impact of high aptitude on performance. Linking rewards to performance should increase motivation and thus the performance of Army personnel.

References

Campbell, J. P., Motivation theory in industrial organizational psychology. In Marvin D. Bunnette (Ed) Handbook of Industrial and Organizational Psychology, Chicago, Rand McNally Publishing Company, 1976.

Ghiselli, E. E., The validity of aptitude tests in personnel selection. Personnel Psychology, 1973, 461-477.

Miller, L., Behavior Management, New York: John Wiley & Sons, 1978.

Prichard, R. D., Holenback, J. & De Leo, P. J. The effects of continuous and partial schedules of reinforcement on effort, performance and satisfaction. Organizational Behavior and Human Performance, 1980, 25, 336-353.

Scribner, B. L., Smith, A. D., Baldwin, R. H., and Phillips. R. L. Are smart tankers better tankers: AFQT and military productivity. West Point, NY. Office of Economic and Manpower Analysis, United States Military Academy, December 1984.

Stouffer, S. A., Lumsdaine, A. A., & Lumsdaine, M. H. Attitudes before combat and behavior in combat. In S. A. Stouffer, A. A. Lumsdaine, M. H. Lumsdaine, R. M. Williams, Jr., B. Smith, I. L. Janis, S. A. Stair, and L. S. Cottrell, Jr. The American soldier: Combat and its aftermath, Vol II of Studies in Social Psychology in World War II, Princeton, NJ: Princeton University Press, 1949.

Tziner, A., & Eden, D. Effects of crew composition on crew performance: Does the whole equal the sum of its parts? Journal of Applied Psychology, 1985, 70, 85-93.

Stress and Aircraft Maintenance Performance in a Combat Environment[1]

William D. Kane, Jr.
Western Carolina University

Abstract

This effort examines the impact of stress on the performance of aircraft maintenance in a combat environment. The theoretical literature and research and cases from actual combat situations indicate that performance could be significantly degraded. The literature is not clear on a definition of stress and no practical measure of stress exists. Also, the relationship between stress and performance is not accurately predictable. However, maintenance capability will probably be degraded by psychological casualties and the intensity of the battle will dictate the psychological casualty rate. The shape of the relationship between combat stress and performance is suggested and recommendations are made as to what additional research might be conducted.

Introduction

Data from World War II and more recent conflicts in the Middle East indicate that in active combat a unit will sustain one psychological casualty for every three physically wounded casualties (Hoey, 1984). In sustained combat, 25% of all casualties may be combat fatigue casualties resulting from an individual's temporary inability to cope with the extreme stress. These data apply to recent conflicts engaged in by combat experienced, well trained soldiers. We can only speculate what the psychological casualty rate might be among Air Force maintenance personnel who have no history, role models, experience, or training with launching, servicing and repairing aircraft in an active combat situation. Since psychological casualties are generally highest in the early stages of the conflict (Hoey, 1984), and we may bring the troops to the conflict already fatigued, we could estimate that the percentage of psychological casualties will be quite high. As Dr. Jones comments "In one battle area during Israel's 1973 Yom Kipper War against Egypt, 900 of the first 1,500 Israeli casualties were psychiatric casualties" (Hoey, 1984, p. 33). A maintenance unit's sustainability could be badly degraded by stress casualties, and since there is no way to predict the distribution across AFSC's, its capability could be completely lost in some areas.

"It has been known for a considerable period that soldiers in the heat of battle are unlikely to fire their weapons. After the Battle of Gettysburg in the American Civil War, over 200 of the muzzleloading rifles were found to have been loaded five or more times without being fired, and one had been loaded 21 times without being fired once." (Idzikowski and Baddely, 1983, p. 128, referencing Walker and Burkhardt, 1965.) Work by Marshall and colleagues in World War II indicated that "...only 15-25 percent of the soldiers involved in an engagement actually fired their weapons." (Idzikowski and Baddely, 1983, p. 126.) Other World War II research indicated that navigational errors

---

[1] This research was sponsored by the Air Force Office of Scientific Research through its Summer Faculty Research Program. The contractor was the Southeastern Center for Electrical Engineering Education and the site was the Air Force Human Resources Laboratory, Logistics and Human Factors Division, Combat Logistics Branch. Contract number F49620-82-C-0035.

increased the closer the bomber got to the target area yet decreased after leaving the target area and heading safely home. Again referencing Walker and Burkhardt (1965), Idzikowski and Baddely (1983) describe research that compared the ratio of error in combat to error in training. "The results show a decrement of up to 900 percent as the combat situation becomes more and more dangerous." (p. 129) After reviewing other studies concerned with the relationship between fear and performance Idzikowski and Baddely (1983, pp. 140-141) conclude "Behaviorally, deterioration can be expected in manual dexterity, in sensory motor tasks such as tracking, and in performance of secondary tasks. It is probable that secondary task performance is reduced before central tasks are affected... The evidence suggests that when a situation has induced fear in an individual... then a deterioration in the efficiency of performance can be expected, especially in tasks involving sensory-motor skill or divided attention."

If we apply the above conclusions to aircraft maintenance in an active combat environment we can speculate about task performance outcomes. As the intensity of the battle increases individuals' fear/anxiety will increase. As fear/anxiety increases task performance degrades with peripheral tasks being neglected first. We can expect increased mistakes in task accomplishment with decreasing attention paid to secondary tasks such as safety. (It might be possible for safety to become the primary task and the worker forgets about the maintenance task!) At some point <u>all</u> tasks are subordinated to survival. The stress/threat of combat <u>will</u> impact on a maintenance unit's performance.

<u>Stress, What is it?</u>

Selye (1979, p. 12) defines stress as "the non-specific response of the body to any demand." This definition dates back to 1936 and is still reasonably valid, if not very helpful. Non-specific response refers to the fact that the sterotypical stress response can be elicited by any number of stressors, rather than specific stressors. Any stressor, therefore non-specific, produces the stress response. The stress response that Selye outlined and labeled the General Adaptation Syndrom (G.A.S.), emphasizes the evolution of stress in three stages.

1. Alarm reaction. This occurs upon sudden exposure to noxious stimuli to which the organism is not adapted. The reaction has two phases:
   (a) Shock phase, the initial and immediate reaction to the noxious agent. Various signs of injury such as tachycardia, loss of muscle tone, depressed temperature and blood pressure are characteristic symptoms.
   (b) Countershock phase, a rebound phase marked by the mobilization of defensive forces. This phase merges into the next defensive phase, during which the adrenal cortex is enlarged and secretion of adrenocorticoid hormones is increased.
2. Stage of resistance. This is marked by full adaptation to the stressor during which symptoms improve or disappear. There is, however, a concurrent decrease in resistance to other stimuli.
3. Stage of exhaustion. Since adaptability is finite, exhaustion inexorably follows if the stressor is sufficiently severe and applied for a prolonged period of time. Symptoms reappear and if stress continues unabated, death ensues. (Selye, 1979, p. 17)

The organism is stimulated, it responds to and adapts to the stressor, and if the stressor is not somehow relieved the organism eventually fails.

In spite of years of work on stress there is currently no agreed upon definition of stress and science is still attempting to fathom its mysteries. Actually, what has been labeled stress should more appropriately be labeled arousal, the result of stress. "The identity of the alarm signals that first relay the stress message has yet to be identified." (Selye, 1980, p. 130)

The first mediator is still unknown and what has been observed and measured is the impact of the first mediator. In addition to Selye's definition, among the many others are "...stress may be considered as a response to a stressor that induced a change in the individual's ongoing behavioral, physiological or cognitive patterns of functioning..." (Beech et al. 1982, p. 10). Another is "Stress will arise whenever the effort mechanism is either seriously overloaded over time or falls altogether short in accomplishing the necessary energetical adjustments." (Sanders, 1983, p. 79). Robins et al. (1961) point out a variety of definitions all stemming from the particular approach, discipline, or philosophy employed. However, in order to proceed, this paper requires an operational definition of stress in spite of the fact that it may be imperfect, incomplete, or even ultimately false. For the purpose of this paper's focus, the Air Force maintainer under combat conditions, stress is that level of arousal, whatever the stressor, that influences performance.

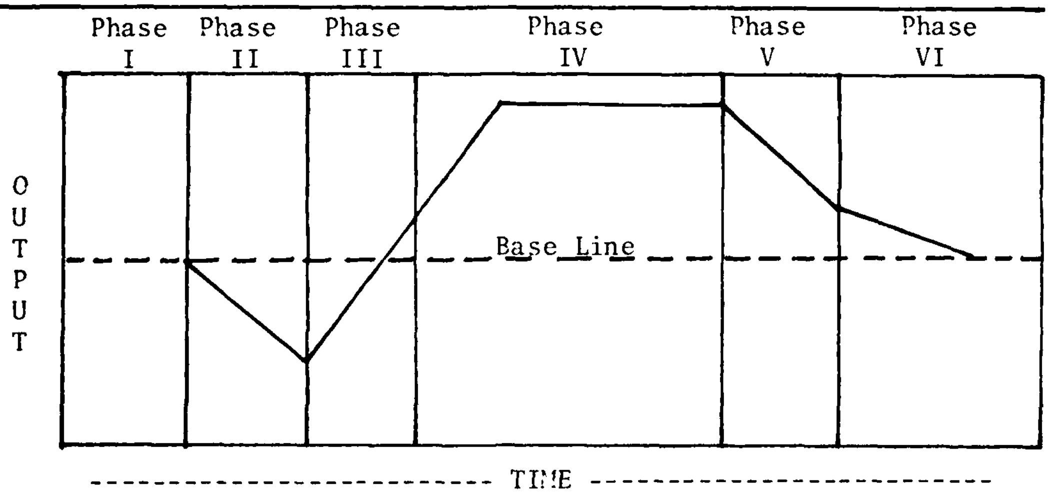
The issue of performance raises another cloud. The relationship between performance and stress is not well understood and most of the research that has been accomplished is simplistic. (One exception is Ursin et al., 1978) Much of the work has been done with one variable, or very few variables, and most of it in the laboratory. The impact of assumed stressors (e.g., temperature, fatigue, noise) is very complex and any generalization from a stressor to performance is premature. Probably one of the most significant stressors is fatigue, and there is reasonable general knowledge on the limits of fatigue. Yet, the evidence indicates that you cannot predict individual performance because the impact of the fatigue stressor may be moderated positively or negatively by a host of other variables. Stress at the extreme this paper is concerned with, under life and death conditions, is suspected of degrading performance, but to what extent and how soon and for which people is unknown. Maintenance troops have multiple stressors, such as life stress, task demand stress, organizational stress, and combat stress, and at the point of interest, performance on the flight line under attack, prediction to or measurement of individual performance is not practical.

Measurement of stress raises other issues. As indicated earlier, what has been measured is arousal and it has been measured in a variety of ways. Unfortunately, the variety of measures produced conflicting results because different stressors cause different reactions and the different reactions measured by different methodologies produce conflicting results. Stress has been measured by self-reports, other paper and pencils instruments, Rorshach tests, urinalysis (catecholamines), blood tests (cortisol, androgens), heart rate, galvanic skin response (GSR), performance, brainwaves, and observation. While there has been good experimental work (e.g., Ursin et al., 1978 and Bourne, 1969) there is not any useful practical work. Given the variety of stressor sources impacting on the individual there is no way to allocate a portion of the total stress to any stressor or class of stressors as there is no valid or practical way to measure. We can assume that in a combat situation the individual's stress will be very high and that most of his attention will be riveted on that which threatens to destroy him. He may or may not have time (attention) to spend on a maintenance task and if he begins the task he may or may not make errors ranging from trivial to dangerous. But, there currently exists no practical way to accurately measure what is going on or why.

Performance in Combat

Cowings (1975) reported on two small Army maintenance support units, one repairing vehicles and one repairing aircraft. Both units experienced unexpected enemy attacks, one including a ground attack, and the attacks had a significant impact on maintenance output (see Figure 1). The initial impact, Phase II, was a decline in performance due to the high state of arousal

generated by the shock of the attacks. In Phase III, output increased rapidly past original base line output, Phase I, to an all time high as the troops



*Phase I Pre-Combat Period
*Phase II Initial Impact Period
*Phase III Recovery Period
*Phase IV Hyper-Efficiency Period
*Phase V Deceleration Period
*Phase VI Post-Combat Period

Figure 1
Impact of Enemy Attack on Maintenance Output
(Cowings, 1975, p. 90)

adapted to the stressful combat environment. In the hyper-efficiency phase, Phase IV, output continues at an abnormal pace as the maintenance troops live off the high stemming from the elevated state of arousal. This period of abnormally high output lasts for some period of time, five to six days in these two cases, and then performance falls abruptly as exhaustion sets in, Phase V. In both of these cases the commanders immediately recognized the need for rest, battle conditions permitted it, everyone took a day off and had a beer party, and output resumed the base line rate, Phase VI. If rest had not been taken it is likely that performance would have continued to decline precipitously.

Continuing to work from Swank and Marchand's (1946) accounts, Kern (1966) reports their scheme for explaining the development of combat fatigue. The stages are: Phase 1, Initial Combat Adaptation; Phase 2, Period of Maximum Effectiveness; Phase 3, Hyperreactive Phase; and Phase 4, Emotional Exhaustion Phase. Note the similarity to Cowings' (1979) conceptualization (Figure 1) (Cowings does not reference Swank and Marchand) and also note the similarity to Selye's (1979, p. 17) development of the stages of the G.A.S. earlier in these pages. Selye's (1979) three stages are: 1. Alarm reaction; 2. Stage of resistance; and 3. State of exhaustion. While the particulars are not very clear, it is apparent that there is agreement that the impact of combat stress is going to shape the output of a maintenance unit. What we do not know at the present is what that impact will be, who it will impact, and to what extent it will impact. We do know that it could be severe and therefore devastating.

Recommendations

There are five recommendations, listed in the order in which they should be undertaken. The criterion for the sequence of the list is the time required before something useful can be produced that will enhance maintenance

performance in combat. The five recommendations are:

1. Develop two education programs for stress management in the Air Force maintenance environment with emphasis on reducing the impact of stress on combat maintenance. One program to focus on all supervisory personnel and one program to begin to condition maintenance personnel for functioning in a combat environment.
2. Conduct a study of Air Force maintenance personnel, in sortie producing AFS's, comparing them against all other Air Force personnel to determine if there is a higher incidence of stress related disease among maintenance types. This would be designed to determine in a macro fashion if maintenance is a more stressful environment.
3. Devise an Index of Organizational Stress. It might be possible to develop a series of indicators that would make up an index of organizational stress that would provide a gross measure of a wing's cumulative stress.
4. Investigate the possibility of teaching individuals better coping strategies whereby they conduct more competent transactions with their environments.
5. Begin experimental work on the flight line to develop multiple measures of stress. Also investigate the relationship of stress to performance in the complex context in which maintenance takes place.

Conclusion

Is the stress of combat going to impact on sortie generation rates? Probably, but the exact impact is not known at this time. Limited data indicate that combat max efforts will only last about five days. Dr. David Jones concurs with this assessment, but comments that it is more of a gut feel than objective data. After that time period exhaustion sets in and productivity declines sharply. Combat stress will impact maintenance performance incrementally as well as catastrophically. Mistakes will be made, tasks will take longer, and some tasks may be ignored or forgotten as individuals concentrate on personal safety. The exact impact of combat stress on maintenance performance is unknown, but it is likely to be significant.

Can anything be done to moderate the impact of combat stress on maintenance productivity? Perhaps. Teaching Air Force supervisors to identify stress symptoms and then engaging in appropriate intervention techniques could reduce the negative impact of stress. Teaching individuals better coping skills could provide them with additional psychological strength to resist performance degradation due to combat stress. Managing stress levels in peace time so as not to bring troops into combat already well along the stress-performance curve could enhance sustainability. The matter of what can be done to moderate the impact of combat stress on maintenance performance is a research question and one that needs immediate attention.

References

Due to space limitations, references are not included. For a complete list of references, contact the author.

Neuroelectric and Neuromagnetic Recordings: Possible New Predictors of On-Job Performance

Lewis, G. W, Trejo, L. J., Blackburn, M. R., and Blankenship, M. H.
Navy Personnel Research and Development Center
San Diego, California 92152-6800

ABSTRACT

Neuroelectric (evoked potential, EP) and neuromagnetic (evoked field, EF) data were obtained from a group of Marine Corps security personnel. HIGH and LOW groups were determined from on-job performance ratings obtained after the EP and EF recording sessions. Higher EP and EF amplitudes were seen for the HIGH group than for the LOW group. Group differences were found for EP recordings over the left hemisphere visual reception area, while greater group differences were found with EF recordings over the right hemisphere visual reception area. No group differences were found with paper and pencil test scores.

INTRODUCTION

Personnel assessment in military and civilian communities has depended heavily on paper and pencil tests. These tests are able to predict academic performance fairly well, but have been criticized for their ineffectiveness in predicting on-job performance (Ghiselli, 1966). New kinds of tests are required to provide more complete understanding of the unique capabilities of individuals. Research on brain function that emphasizes "process" rather than "content" variables, may be able to predict on-job performance better than the conventional tests. This Center has been investigating the use of neuroelectric (evoked potential, EP) recordings to assess individual differences in brain processing and their relationships to on-job performance. Temporal stability is a prerequisite for using EP data for personnel assessment and has been shown in this and other laboratories (Lewis, 1984). Relationships have been found between EPs and on-job performance in pilots/radar intercept officers (F-4 fighter aircraft), antisubmarine warfare trainees (sonar simulator) and enlistees (promotions and attrition). Research to date has suggested that EP records are better able to discriminate and classify on-job performance groups (e.g., high vs. low) than conventional paper and pencil test scores. Another consistent finding from this laboratory is that EP data are more variable from low performing than from high performing personnel (Lewis, 1983).

For the last four years, this Center has been developing neuromagnetic (evoked fields, EF) recording capability which has several advantages over traditional neuroelectric procedures. These include being noncontact, monopolar, showing increased sensitivity and localization of brain activity, providing new information, and possibly increasing capability to assess individual differences, and therefore predicting on-job performance (Lewis, 1983; Lewis & Blackburn, 1984). We have reported EF single epoch recordings for the first time and EF reliability (Lewis, Blackburn, Naitoh, & Metcalfe, 1985).

In this paper, we describe our field-recorded EP and EF data and their relationships to on-job performance.

METHOD

EP, EF and on-job performance data were obtained from 26 Marine Corps security guards. On-job performance data dealt with military and job knowledge and performance, reliability and motivation. Two groups (HIGH and LOW) were determined from these criteria data. The mean age for the HIGH group (N=16) was 21 ± 2 years, while that for the LOW group (N=10) was 20 ± 1 years. Data were also obtained from two paper and pencil tests, the Cognitive Laterality Battery (CLB) (Gordon, 1983) and the Test of Attention and Interpersonal Style (TAIS) (Nideffer, 1977). The CLB is a series of tests that assess cognitive functions such as verbal/sequential and visuospatial processing. It has been used in diverse occupational groups as combat pilot trainees, bank employees, and computer programmers. The TAIS is a self-report inventory that assesses the respondent's ability to control attention and interpersonal factors. Such factors have been suggested as important in emergency situations, competitive athletics and business. This test has been used as a personnel selection battery which includes police applicants.

Each subject viewed (binocular, central fixation) a black and white checkerboard pattern subtending 5 degrees visual angle (DVA) at a luminance of about 10 ftL. Each check subtended 0.4 DVA. The stimulus was flashed on for 10 msec. Interflash interval varied between 500 and 1500 msec. Background luminance was about 1 ftL. EP data were recorded using a commercially available electrode helmet (Electro-Cap International), amplified (20,000 gain) and bandpassed (0.1-100 Hz; Grass amplifiers, model 12A5). Ten channels of data were obtained, however, data from only two sites (i.e., visual reception/occipital area 01,02) will be reported in this paper. EF recordings were obtained using a DC SQUID Biomagnetic Detection System (B.T.I., Inc., model 600B, second derivative gradiometer). The single channel EF signal (1000 gain on SQUID control unit) was bandpassed (0.1-40 Hz Krohn-Hite, model 3343) and further amplified (50 gain, Grass P511J) prior to digital conversion. Sampling rate for the EP and EF recording was 256 Hz. Poststimulus record lengths were one-half second. Due to artifact rejection, EPs were averaged over 7 epochs while EFs were averaged over 19 epochs. All EP and EF data were acquired and stored as single epochs on a field-portable computer system (MASSCOMP, model MCS-560). The unit of measure for the EPs was the microvolt (uV) while that for the EF was the femtoTesla (fT) (10^{-15} Tesla). Sample EP and EF data appear in Figure 1 recorded over the left (01) and right occipital (02) areas. Note the similarity in the EP data recorded over the two separate areas (01 vs. 02) and the polarity reversal in the EF data recorded from the same general areas (01 vs. 02). More precisely, the EF data were recorded 1 cm lateral to the 01 and 02 EP sites. However for convenience, the EF site locations will be referred to as 01 and 02. For both the EP and EF records, root mean square (rms) amplitudes were obtained from each single epoch and the averaged data. Due to large individual differences in EP and EF records, it has been found that by using an integrated amplitude measure (e.g., rms), at least one number per waveform may be obtained where particular waveform components may not be well defined (Lewis & Froning, 1981).

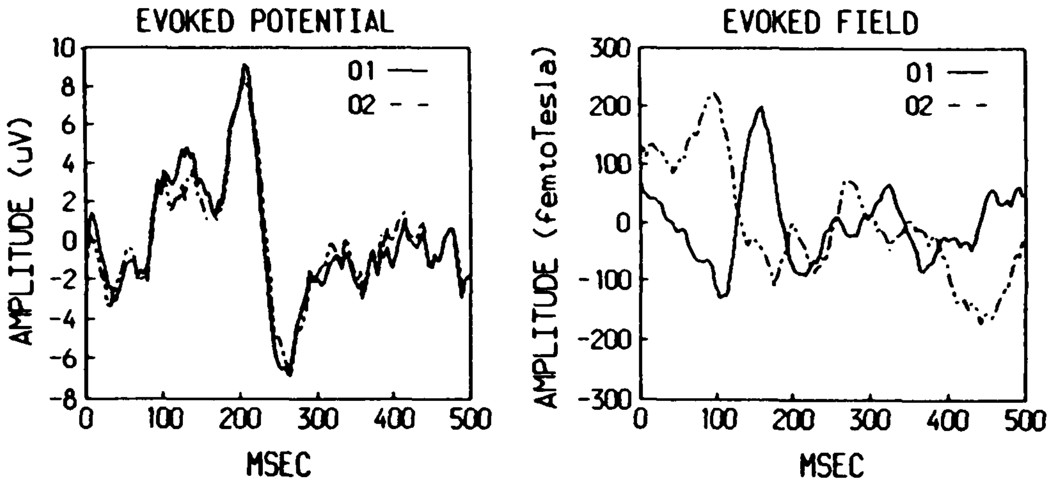


Figure 1. Sample EP and EF data.

RESULTS

Paper and Pencil Tests

Neither the TAIS nor the CLB scores were able to distinguish the HIGH group from the LOW group (t-test). The CLB did show that both groups of security personnel performed better on tests of visuospatial function than on verbal/sequential tests (nonparametric sign tests, $p < .02$). This finding suggested that the subjects were a homogeneous sample based on the CLB.

Evoked Potential and Field Data

Mean, standard deviation, coefficient of variation (CV=SD/MN) and t-test data for EP and EF recordings, sites and performance groups appear in Table 1. Three LOW group and two HIGH group subjects lacked EP data reducing the group sizes to N=7 and N=14 for the LOW and HIGH groups, respectively. The LOW group had lower amplitudes than did the HIGH group for EP and EF recordings over both sites (01 and 02). Evoked potential CVs were about the same for both LOW and HIGH groups, however, were much greater for the LOW than for the HIGH group a both sites for EF recordings. Group differences were found for the EP data at site 01 ($p < .05$) and for the EF at site 02 ($p < .003$). Even though the EF SDs were about the same for both groups at site 02, the mean value for the HIGH group was nearly two times that for the LOW group (Table 1, Figure 2). Largest group differences were seen at the EF site 02, which is reflected in the large t-test value and p value in Table 1.

Table 1

Descriptive and Inferential Statistics for Evoked Potentials and Fields, Sites, and Performance Groups

	Evoked Potentials (μ Vrms)				Evoked Fields (fTrms)			
	Site 01		Site 02		Site 01		Site 02	
	Low (N=7)	High (N=14)	Low (N=7)	High (N=14)	Low (N=10)	High (N=16)	Low (N=10)	High (N=16)
Mean	5.00	6.99	4.83	6.73	217	272	173	331
STDEV	1.70	2.20	1.71	2.47	145	105	115	119
CV	.34	.31	.35	.37	.67	.39	.66	.36
t	2.08		1.81		1.12		3.33	
df	19		19		24		24	
p	.05		.08		.27		.003	

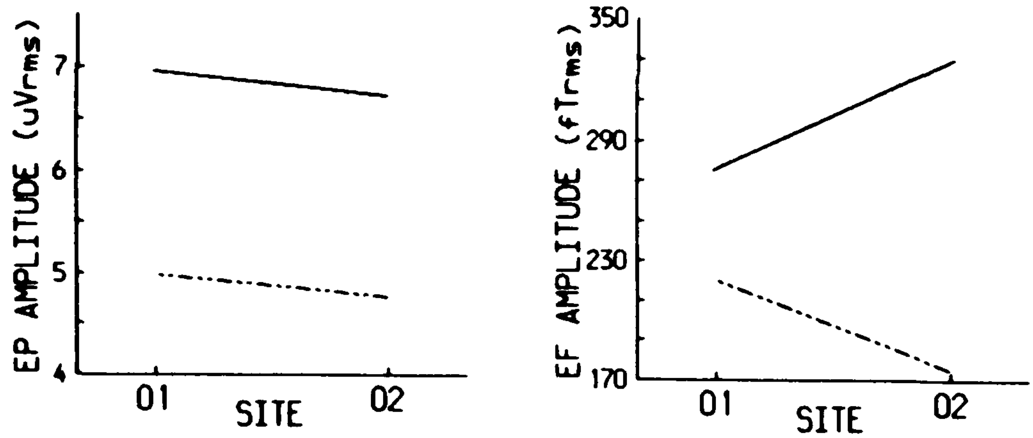


Figure 2. Group mean values for EP and EF data at sites 01 and 02. HIGH group drawn in solid lines, LOW group in dashed lines.

DISCUSSION

To our knowledge, the data presented in this paper are the first known neuromagnetic EF data to be acquired outside of a highly controlled laboratory setting. Magnetic noise was recorded from this operational environment and are undergoing analyses. Noise sources such as floor buffers, soft drink machines, and two-way radios were picked up occasionally by the neuromagnetometer. When magnetic artifacts were present, the records were rejected on-line.

In our laboratory, we have consistently seen greater intra- and inter-individual variability in low performers than in high performers. The coefficients of variation of the EF rms amplitudes presented in Table 1 also support this generalization. Inter-group variability in the EP data was not as pronounced as for the EF data. Improved localization of EF recording over EP recording may account for the large site and group differences. These neuroelectric and neuromagnetic findings showed group differences that were not seen in the CLB and TAIS data. In contrast, the CLB and TAIS tests suggested behavioral homogeneity among the subjects and did not correlate with job performance ratings.

Other research projects continue to explore the feasibility of using neuroelectric and neuromagnetic measures to assess baseline and high stress performance conditions and reduce training attrition of naval aviators.

REFERENCES

Ghiselli, E. E. (1966). The validity of occupational aptitude tests. New York: John Wiley.

Gordon, H. W. (1983). The cognitive laterality battery. Produced by Media Productions Services. Pittsburgh, PA: Western Psychiatric Institute and Clinic, University of Pittsburgh.

Lewis, G. W. (1983). Event related brain electrical and magnetic activity: Toward predicting on-job performance. International Journal of Neuroscience, 18, 159-182.

Lewis, G. W. (1984) Temporal stability of multichannel, multimodal ERP recordings. International Journal of Neuroscience, 25, 131-144.

Lewis, G. W., & Blackburn, M. R. (1984). Biomagnetism: Possible new predictor of personnel performance. (NPRDC Tech. Rep. 84-43). San Diego, CA: Navy Personnel Research and Development Center.

Lewis, G. W., & Froning, J. N. (1981). Sensory interaction, brain activity, and reading ability in young adults. International Journal of Neuroscience, 15, 128-140.

Lewis, G. W., Blackburn, M. R., Naitoh, P., & Metcalfe, M. (1985). Few-trial evoked field stability using the DC SQUID. In H. Weinberg, G. Stroink, & T. Katila (Eds.) Biomagnetism: Applications and Theory. New York: Pergamon Press, pp. 343-347.

Nideffer, R. M. (1977). Test of attentional and interpersonal style. San Diego, CA: Enhanced Performance Associates.

Partially funded by the Defense Nuclear Agency. The views expressed in this paper are those of the authors and are not necessarily the views of the Department of the Navy or Defense Nuclear Agency. Appreciation is expressed to Ms. Brandi Jackson for collecting the CLB and TAIS data and Mr. Paul DeYoung for collecting and analyzing the CLB and TAIS data.

DEVELOPMENT OF COMPUTER-BASED MAINTENANCE AIDS SYSTEM

JEFFERY D. CLAY
AIR FORCE HUMAN RESOURCES LABORATORY

ABSTRACT

To accomodate the rapidly growing volume of technical data required to support modern weapon systems the Air Force is rapidly moving toward the digital storage and presentation of Technical Orders (TOs). The Air Force Human Resources Laboratory (AFHRL) believes that the maintenance technician's needs should be considered first in the design of such a system. AFHRL has done extensive human factors research into the format, man-machine interface, and data content problems of displaying TO data on a computer. A prototype system was developed to test data presentation and man-machine interface techniques. This system was evaluated in an intermediate-level shop at Grissom AFB IN. The evaluation was accomplished by having technicians perform maintenance tasks using the data presented on the system. Data was collected on the technician's task times and attitudes about the new system. The results will provide the basis for a specification for use in procuring automated TO data.

With the rapid growth of information needed to support and maintain modern weapon systems the Air Force is moving towards the automation of Technical Orders (TOs). This automation has the potential to save both time and money in initial distribution of TO's and in the distribution of updates. The Air Force Human Resources Laboratory (AFHRL) believes that the performance of the maintenance technician can also be improved through quicker access to necessary information, the use of information that is more complete and tailored to the technician's skill level, and the increased accuracy of the information. To insure that the technician's needs were incorporated into the design of such a system AFHRL initiated a project to study these needs to develop a prototype system and to develop a specification for the procurement of automated information systems. This project is known as the Computer-based Maintenance Aids System.

Several design studies using Air Force maintenance technicians were done to determine the optimum methods for displaying complex text and graphic information. This information was initially tested in a field test developed by Hughes Aircraft Corp at Offutt AFB NE. Although testing of the system provided many valuable insights, the system contained several design features which resulted in it not being accepted by the user. Using the lessons learned from Offutt, a second prototype

system was developed in-house by AFHRL personnel and tested at Grissom AFB IN. This prototype and field test proved successful and provided the data required to produce a specification on how to present automated technical information.

The first prototype designed by Hughes was installed in the 55th AMS Shop at Offutt AFB during Nov-Dec 84. While this test demonstrated the feasibility of automated TO's, there were several factors which made it unacceptable. Among these were the slow system response time, some graphics problems, and multiple programming errors. The slow response time between frames of information was the major factor affecting user acceptance. The time between selecting a choice and having it displayed averaged about 10 seconds with some frames requiring as long as 30 seconds. The technicians were unwilling to sit and wait this long for a response and therefore did not want to use the system when they could use a paper TO faster. This problem and the graphics and programming problems greatly inhibited user acceptance of the system. While marring the results of this field test, these were important lessons learned and applied to increase the user acceptance of the second field test.

The Grid Compass II model 1139 computer was selected as the hardware for the second prototype. The Grid was chosen for its small size and its powerful capabilities which made it an ideal candidate for such a test. The data used for this field test applied to the RT-728A/APX-64 radio receiver-transmitter. The Checkout and Analysis section of the data was chosen and analyzed to determine any additional sections needed to support the checkout. These additional sections included parts of: Theory of Operation, Illustrated Parts Breakdown, and Troubleshooting. Additional troubleshooting routines were developed by an experienced technician at Offutt AFB. These routines provided for troubleshooting one of the modules of the RT unit.

The software for the Grid was developed to store and present TO data on the computer. This software provided the necessary functions such as scrolling, direct access, logical branching, and direct keyboard input into the program for technician interaction. Scrolling allows the technician to manipulate a drawing which is too big to fit on the screen such as a schematic or other fold out graphics. Direct access allows a technician who is experienced with the system to bypass the menus and go directly to the desired part of the procedure. Branching allows the computer to display only the information needed by the technician when two separate paths exist (such as troubleshooting). This limits the amount of irrelevant information the technician must look at during a procedure.

The data was reformatted into an integrated text-graphics style as shown in figure 1. This format is similar to that used for job guide manuals. The top and bottom of the screen were reserved for information telling the technician which section of the TO he was in and the options available to him. The middle part of the screen was divided into sections with the left two thirds reserved for text and the right third reserved for graphics.

The header section included the TO number, the name of the section, and the number for the particular frame. This information was used for direct access. The information displayed at the bottom of the screen included the different options available from that frame. These included Theory of Operation, Schematics, Direct Access, Illustrated Parts

Breakdown, Table of contents, or other sections which related to that particular frame. The Return option returned the user to the point in another procedure where he requested other information. An example of this would be returning to the Checkout after paging through several frames of Theory of Operation to learn more about the operation of the radio.

The space on the left side of the screen is reserved for text with the associated graphics displayed on the right side of the screen. The text is single-spaced for procedures with double-spacing between steps. If a drawing is larger than the screen, scrolling is implemented within as large a window as possible with text, if needed, appearing outside the window. When scrolling is possible an icon consisting of four arrows pointing up, down, left, and right is displayed in the bottom right corner of the screen.

The procedures were written in two levels of detail: Track One for the experienced technician and Track Two for the inexperienced technician. Track One was made up of only a checklist of steps necessary to accomplish a task, which is adequate for an experienced technician. Only the necessary graphics accompanied the text, such as wave forms or disassembly drawings when needed. Track Two included more detailed text on how to accomplish steps along with graphics depicting where the required operation was to take place. Some procedures required more frames in Track Two than the equivalent information in Tack One. The user could select the desired level of detail and always had the option to switch tracks at any time. Support information such as Theory of Operation was made up of only one level of detail.

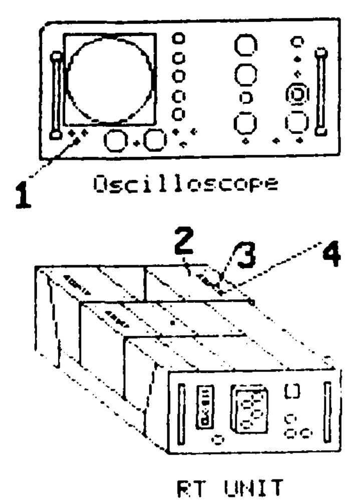
CMAS II was taken to an intermediate level shop at Grissom AFB IN during August 1985 for a two week field test to determine the feasibility and user acceptability of an automated TO. A short training session was

12P4-2APX64-2 PRELIMINARY OPERATIONAL CHECK 2.2.5.1C

g. Connect a 1:1 scope probe from Channel A of AN/UPM-137A oscilloscope VERTICAL CHAN A VIDEO IN jack (1) to each test point on A8 module (2).

h. Verify that Test Point A8TP1 (4) has a reading of 0.05 VP-P MAX

i. Verify that Test Point A8TP2 (3) has a reading of 0.05 VP-P MAX



NEXT = SPACE BAR BACK = B OPTIONS = O

FIGURE 1

given to all technicians before the test began and each technician had to perform a simple test which proved they could use all options available in CMAS II. Technicians of various experience levels were asked to use CMAS II as their TO to perform the checkout and analysis of the receiver-transmitter. Known faults were introduced into the radio and times were recorded for the technicians to find the fault. A similar task was then performed using the paper TO. Because of limited personnel and the necessity to have the technicians available for other tasks, the field test was conducted on a non-interference basis. Eight technicians were able to complete all phases of the test. After completing the test, the technicians were asked to complete a questionnaire giving their reactions to CMAS II and suggestions for improvements. The Technicians were also interviewed to obtain other comments.

The overall response to CMAS II was very favorable. The feature which received most of the negative comments was the schematic diagram. This was due to the small size of the screen and the limited view of the schematic which could be seen at one time. All other aspects of the system were readily accepted by the technicians. Here are some of their comments:

> "The computer is so easy to use. It formats things and limits the amount of information that you see at a time which limits the confusion. It puts information in a step-by-step format which is easy to follow."

> "At first I didn't like the schematics because they show such a limited view but after a little familiarization it seems that they may not be as good as the TO but they are sufficient."

> "I liked having both levels of detail. I used both levels but preferred the more detailed level. I liked the illustrations too. They tell you exactly where to put the connections. They are especially good for people who have not gone through the procedure before."

> "When you go through the checkout you must read every step. It is easy to skip over a step in the TO. With the computer you are not as likely to skip over a step."

> "I would like to see the computer implemented in the Air Force. I thought it was a joke at first but now I think it will work. I liked using it."

> "I like it! I have a very positive reaction to it. You can get information a lot faster once you become acquainted with the system. It sure beats flipping through the pages of a TO."

> "I really like the way the IPB (Illustrated Parts Breakdown) works. I know how much time it would save me because it is my job to order parts."

The results obtained from these field tests will be incorporated into draft specifications, outlining the hardware capabilities and the technical data requirements necessary for effective automation of technical information. These specifications will be available to Air Force or other DOD agencies interested in the acquisition of automated TO's.

Automated technical information is not only desirable and helpful for the maintenance technician, but will be essential to keep pace with the weapons systems of the future. This technical information must be changed from today's format to be interactive with the technician and provide some support in troubleshooting the complex electronics of tomorrow's airplanes. AFHRL believes that an automated TO and information delivery system should be capable of not only supplying technical information, but also retrieving managerial and diagnostic information from other available sources. These sources would include both ground-based and airborne information systems which are now under development. To insure that such a system is useful in keeping the United States Air Force the best flying force in the world, the technician's needs and capabilities must remain foremost in the design and implementation of such a system.

EVALUATION OF WALTER REED PERFORMANCE ASSESSMENT BATTERY
IN REPEATED MEASURES RESEARCH

Ronald J. Edwards and George S. Siering*
U.S. Army Aeromedical Research Laboratory

ABSTRACT

The U.S. Army Aeromedical Research Laboratory (USAARL) has conducted repeated measures research using versions of the Performance Assessment Battery (PAB) developed by the Walter Reed Army Institute of Research. This paper is the first report evaluating the stability of the PAB as implemented at USAARL. Thirteen volunteer Army aviators took the PAB ten times over four days. Means, standard deviations, and cross session correlations were calculated and examined. In general, the PAB has three good tests, three which are marginal, one which is slightly unstable and one which is not satisfactory in its present form. Most PAB subtests lack initial stability and pre-experimental training will be required to reduce the practice effects.

INTRODUCTION:

The Performance Assessment Battery (PAB) is a computerized test battery developed by the Walter Reed Army Institute of Research (WRAIR). It is composed of a variety of subtests which measure varying degrees of cognitive and visual-motor processing abilities. To date, it has been used in experiments studying the effects of sleep deprivation, sustained performance, jet lag, physical fatigue, hypoxia, sickle cell disorders (Thorne,1982), heatstress (Hamilton, 1983; Mitchell,1985) and atropine effects (Simmons,1984). This study is designed to evaluate PAB for use in assessing the effects of a stressor on the human system by repeated testing of unstressed subjects.

The two primary research questions asked were: 1) How do PAB results change when a given subject is tested repeatedly (learning curve) or, asked a different way: How many times need a subject take the PAB test before further change in scores can be assumed due to some imposed stressor of research interest rather than a learning curve? 2) How do PAB test results differ between subjects, given the same exposure to the test routine?

METHOD

SUBJECTS:

The 13 subjects were newly-rated Army aviators who had recently completed flight training at Fort Rucker, Alabama, and were awaiting the beginning of their first assignment. All subjects were volunteers and all signed a standard consent form. Their total hours of flight time varied from 160 to 250, and their ages varied from 21 to 30 years. There were 11 males and two females.

APPARATUS:

The version of the battery used in this study included six performance tests and two affective scales. A visual search and recognition task (MAST 6) was used with six target letters at the top of the screen followed by 20 letters in the middle of the screen. The subject determined as quickly as possible whether all target letters were present in the string or not.

Logical reasoning (LOGICAL) was a transformational grammar task, requiring the subject to indicate whether the phrase (i.e., B follows A) correctly described two letters presented on the screen (i.e., AB).

The Probe Memory (PM) task presented nine random digits in a row across the screen for 1 second, followed by a 3-second blank screen, followed by representation of eight of the original nine digits in a different randomorder. The subject had to indicate which of the original nine was missing.

The Serial Addition/Subtraction (SA) task presented a single random digit which further was replaced by a second single random digit which then was replaced by either a plus or a minus sign. The task was to do the indicated arithmetic operation on the two numbers and enter the answer, subtracting 10 if the answer was greater than 10 and adding 10 if the answer was less than 0.

Matrix (MX) was a spatial memory recognition/recall task that presented a random pattern of asterisks on the screen for 1.5 seconds, blanks the screen for 3.5 seconds, then presents a second pattern that the subject determines was either the same or different than the first.

The Wilkinson 4-choice reaction time test (WK) required the subject to observe four small red lights arranged in a square and to sequentially push one of four buttons arranged in a square that corresponded to the position of the lighted light.

Finally, the Mood scale presented the subject with 65 adjectives and collects the five-choice response that best describeed the degree to which that adjective describeed how the subject felt at the time. The Depression (DEP) portion of the Mood test generates a score from 1 to 100 representing the low to high depression range. The Activation (ACT) portion of the Mood test generated a 1 to 100 score representing low to high energy or activity levels felt by the subject.

The origin of each test is given in (Thorne, et al., 1983) and further information is available from WRAIR (WRAMC, Washington, DC 20307).

PROCEDURE:

Each subject had one demonstration session of the PAB test prior to beginning the consecutive 3-day test phase. Specific and consistent instruction was given by the same instructor to each subject. The PAB test then was administered to each subject three times each day (at 0800 hours, 1000 hours, and 1330 hours) for 3 consecutive days. A tenth test session then was accomplished at either 1, 3, 7, or 14 days after the thir test day. The tests were conducted in a soundproof booth in normal room light, temperature, and humidity conditions. The test battery required about 30 minutes per session to complete. The subject was seated in front of a typewriter-style keyboard with a video screen display in front of him. An Apple II+ computer generated the screen display of test instructions and test items. No prior familiarity with computers was required; single numerical key strokes or positional key stroke responses were the required response. The subjects initiated each subtest as they were ready timed only from when the subject self-started the test session. The subjects were not
retained in the test area except for the period of the actual test and no other limitations were placed on activity, diet, sleep patterns, etc.

The data was collected on-line by the Apple II+ computer and was transferred later to a VAX 11/780 for statistical analysis by BMDP software package (1984 version).

ANALYSIS:

The 10 trials on the PAB were analyzed for stability and reliability (cf., e.g., Jones, 1980, Bittner & Carter, 1981). Three measures are available for each of the 6 performance measures employed in this battery. These include Throughput (TP), Speed (SP), and Precent Correct (PC). "...throughput is equal to the number of correct responses, divided by cumulative reaction time....Numerically, the throughput measure is also equal to precent correct, divided by mean reaction time, times a constant. Thus it is a speed-accuracy product." (Thorne et al. 1983, p. 3) Speed was calculated from the total number of responses per minute and may be considered a cumulative reaction time measure. Percent correct, a derived score, was the number of hits divided by number attempted, times 100. These three scoring approaches produce differing outcomes only to the extent that scores depart from 100% correct. However, Throughput scores are recommended by the developers of PAB (Thorne e.al. 1983) and were therefore selected as the measures of choice, for analyses for stability and reliability. Other scores were not evaluated (i.e., Speed and Percent Correct) unless the Throughput scores appeared unsuitable. Mood scores (DEPRESSION and ACTIVATION) were in Percent Correct.

For the present analyses it was considered that tests were stable if: 1) the means over trials exhibited either plateaus, asymptotes or trends where the rate of change of slope was zero; 2) the standard deviations were constant or changed proportional to the mean; and 3) the cross trial correlations were constant. Reliability was considered "excellent" if the average stabilized intratrial correlation was greater than r = .707; "satisfactory" if they were r = .45-.707; unsatisfactory" if less than r = .45.

RESULTS

Analysis of Means

Both memory tasks (MATRIX and PROBE MEMORY) show level Throughput means indicating stability (as well as no improvement) for this metric over trials. Percent correct means for these two tests show learning curves which also appear stable, but offer no other advantages over Throughput. SERIAL ADDITION seems to have fairly level scores after the four initial trials. MAST 6 contains anomalous scores on trial 6 and 9 but also follows an atypical positively accelerated learning curve over other trials. Final evaluation of stability for this test must await analysis of stability of variances and correlations. The two other mean scores available for this test (Speed and Percent Correct) afford no advantages over the Throughput, being similarly afflicted on trials 6 and 9. WILKINSON and LOGICAL REASONING exhibit strong linear trends and WILKINSON shows less signs of slowing reaction times over the 10 trials than LOGICAL REASONING. For both of these tests a transformation may be desirable. Mean ACTIVATION scores are level over nine trials, but those for DEPRESSION decrease. For both of these latter tests scores on trial 10 our our of ordinary bounds.

Analysis of Standard Deviations

All the tests show more day-to-day variability than is desired in a repeated measures application of performance test batteries, but it is to be recalled that these results are based on only 13 cases. None of the tests appears to show decreasing variances, a sign of ceiling effects, or other trends which indicate more critical forms of instability. WILKINSON is the most constant, being level after trial 1, followed by Logical Reason (trial 2), and Serial Addition (trial 3). Mast 6 and Matrix are level but each has two anomalous trials as mentioned above and Probe Memory is quite variable over all trials. The Activation Scale is level over all trials but the last. Depression variances are generally larger than the mean, implying insensitivity rather than instability. Whereas the Speed standard deviation scores for Probe Memory are quite regular, the means offer no advantage over the Throughput measures nor do standard deviations for Speed and Percent Correct scores for the other two questionable tests (Matrix and Mast 6).

Analysis of Correlations

Mast 6 - The intertrial correlations are irregular until session four after which they appear to stabilize with an average value around r = .707 or better.

Logical Reason - The correlations are high (average r more than .83), and appear stable from the first session.

Probe Memory - These correlations may be stable from the beginning, and appear to become stronger after session 4 but remain in the low .60s on the average. The intersession correlations matrix for the Speed measure have similar form and show slightly higher correlations.

Serial Addition - all correlations are high except for those on session trial 10. The remaining values are usually greater than average r = .90 and stability appears to be available from session 1.

Matrix - There is some evidence of superdiagonal form prior to trail 5 or perhaps 6. After trial 6 these correlations appear to stabilize with average values around r = .75.

Wilkinson - There is the suggestion of superdiagonal form throughout this matrix, and while the correlations are at a high level (r more than .85) the slight instability in correlations, plus the rapidly changing means implies that longer practice trials must be given for this test to be used in repeated measures studies of environmental stress.

Depression Scale - These intersession correlations are high, but get higher after trial 4. Trial 10 again has the lowest correlatinos and should be dropped from analysis.

Activation Scale - These correlations are generally low (average r less than .50s) but there is not obvious instability present.

DISCUSSION

In general the PAB has three good tests, (LR, SA, Dep), three which are marginal, (MAST 6, PM, MX), one which is slightly unstable (WK) and one which is not satisfactory ACT) in its present form. The authors conclude from the data that most PAB subtests lack initial stability and that pre-experimental training will be requird to reduce the practice effects. In one current USAARL research program experimenters use three pretest training trials to reduce practice effects in experimental data and this may not be adequate.

We followed the convention of the developers of PAB in using Throughput as the chief dependent variable when evaluating tests for stability and reliability. Some of the reasons for a test's poor showing was insufficient trials or subjects. The best tests were Log Reasoning, Serial Addition, and the Activation scale from the Mood Adjective Check List. We believe that the anomalous scores on Mast 6 and Matrix contributed to our marginal rating for those tests, and perhaps if a larger sample were tested, the irregularities in day-to-day variability would be less bothersome. Probe Memory suffered somewhat from the same problem, but the average retest reliabilities were low enough so that it may be necessary to opt for a longer session when this test is given. The Wilkinson test was marginally unstable, but possessed very high correlations and only slight superdiagonal form. We believe that the differential stability is available from early in practice, but the means increase so markedly over trials, that more practice should be permitted prior to the application of an environmental stress.

For those tests which proved stable, Throughput scores appeared to be at least as good as the alternatives. In those cases where Throughput scores did not meet minimum requirements the alternative scores were not better. We are in favor of continued use of the Throughput score because it has the additional advantage that it is likely to be highly (perhaps even perfectly) correlated with Number Correct. Number Correct is a score directly available from the data and is generally considered to possess sounder statistical and metric properties (Cronbach and Furby, 1970; Bittner, 1979) than any derived (percent, difference, ration, slope) scores. (cf. also Carter, Krause, & Harbeson, 1985).

We would propose that the PAB, with some changes, should be administered to a larger population over perhaps more trials in order to determine better the stability and reliability of the marginal tests and also to examine the richness of the factor composition of the eight tests. We would also propose that the battery be subjected to a task analysis in an effort to link up the constructs measured by this battery with military jobs.

Table I

Stability, reliability, and overall evaluation of PAB Tests administered over 10 trials to 13 Subjects

NAME	STABILITY BY WHICH TRIAL? MEAN	VARIANCE	CORRELATION	RELIABILITY AVERAGE r	OVERALL EVALUATION
Mast 6	anomalous scores	irregular	4	.70s	marginal
Log Reason	Throughput still incr.	very large but level by #2	1	.83s	good (maybe)
Probe Mem	level by #4	irregular	1	.60s	marginal
Matrix	anomalous scores	irregular	6	.75s	marginal
Wilkinson	hyperbolic trend	level by #1	slight unstable	.85s	unstable more prac.
Depression	decrease	irregular	2	.90	good?
Activation	level by #1	level by #1	4	.30	unstable

* The authors wish to acknowledge the review and suggestions by R.S. Kennedy and N.E. Lane of Essex Corp.

CITATIONS:

Bittner, A.C., Jr. (1979). Statistical tests for differential stability. Proceedings of the 23rd Annual Meeting of the Human Factors Society (pp. 541-545). Santa Monica, CA: Human Factors Society. (Research Report No. NBDL-81R010, 10-14). New Orleans, LA: Naval Biodynamics Laboratory (1981). (NTIS No. AD A111086)

Bittner, A.C., Jr., & Carter, R.C. (1981). Repeated measures of human performance: A bag of research tools (Research Report No. NBDL-81R011). New Orleans, LA: Naval Biodynamics Laboratory. (NTIS No. AD A 113954)

Carter, R.C., Krause, M., & Harbeson, M.M. (1985). Beware the reliability of slope scores for individuals. Manuscript submitted for publications.

Hamilton, B., Simmons, R., and Kimball, K., Psychological effects of chemical defense ensemble imposed heat stress on Army aviators. Fort Rucker, AL: US Army Aeromedical Research Laboratory, 1982. USAARL Report No. 83-7.

Jones, M.B. (1980). Stabilization and task definition in a performance test battery (Final Report on Contract No. N0023-79-5089, Monograph No. NBDL-M001). New Orleans, LA: Naval Biodynamics Laboratory. (NTIS No. AD A09987)

Mitchell, G., Knox, F., Edwards, R., Schrimsher, R., Siering, G., Stone, L., Taylor, P., Microclimate cooling and the aircrew chemical defense ensemble. Fort Rucker, AL: US Army Aeromedical Research Laboratory, 1984. In press. .p 5,1,3

Thorne, D., Genser, S., Sing, H., & Hegge, F., (Plumbing Human Performance limits during 72 hours of high task load.) In Forshaw, S.E.(Ed.), Proceedings of the 24th Defense Research Group Seminar on The Human as a Limiting Element in Military Systems, Volume 1. Toronto: Defense Research Group, 1983..p 5,1,3

Thorne, D., Documentation: The Walter Reed Performance Assessment Battery, Walter Reed Army Institute of Research, unpublished, 1982.

SUBORDINATE COMMUNICATIONS IN A MILITARY SETTING

Tim O. Peterson
Texas A&M University

Alan E. Rohrs
United States Air Force Academy

Abstract

The importance of upward communication has been written about and discussed for over two decades. The question of why subordinates are motivated to communicate upward has not. This paper presents empirical findings on motivation and upward communication.

Introduction

In military organizations, the commander is often advised to use an "open door policy." Civilian leaders are encouraged to promote upward communication. Whichever term one chooses to use, the essence is the same, to have a process by which subordinates are systematically encourage to share their feelings, ideas, and concerns with their supervisors. This upward flow of information serves as the invisible web which holds the organization together. It provides both control and coordination for the organization (Poole, 1978). According to many theorists upward communication is both vital (Lippert, 1984; Halatin, 1982; Driver, 1980a; Inman, 1977; Harriman, 1974; Roberts & O'Reilly, 974) and healthy (Hawkins & Preston, 1981; Lahiff & Hatfield, 1978).

One reason to encourage upward communication is that much of the important information a leader needs to make decisions flows up from the organization's lower levels. It is through access to this information that a leader can make decisions which assist in mission accomplishment. Brigadier General Abel (1984), Colonel Strait (1982), and Colonel Winstead (1980) have all written on the importance of maintaining and generating upward communication to help the subordinate understand and appreciate the organization's goals and objectives. They state such communication has a direct relationship to the productivity of the unit. Beck (1985) comments that commanders increasingly find themselves in the position where the superior is managing uncertainty rather than certainty and the subordinate is the expert with information the superior needs. According to the Dana Corporation, "the worker who does the job always knows more about it than his boss" (Peters and Austin, 1985, p. 215).

In addition to productivity, commitment to the organization is another desired outcome of upward communication. Penley and Hawkins (1985) found that career intentions and commitment to the

organization is directly related to supervisor-subordinate communications. A similar finding was been reported by the Leadership and Management Development Center (1980). Brigadier General Abel (1984) states that military leaders cannot motivate, inspire, or persuade their subordinates without a two-way communication flow. It seems then that military leaders must encourage and promote a continuous stream of upward communication from all levels within the organization.

An area requiring research seems to be why subordinates choose to, or are motivated to, communicate upward. This factor is noticeably absent from Glauser's list of influential factors (1984). Over two decades ago, Jackson (1959) noted that the forces that direct communication in an organization are, on the whole, motivational. Employees tend to communicate in order to accomplish some goal, to satisfy some personal need, or to try to improve their immediate circumstances.

In fact, some initial work has been done on the relationship between subordinate motivation and upward information flow. Krivonos (1978) examined the relationship between a subordinate's intrinsic-extrinsic motivation level and his/her perception of satisfaction with upward communication. He hypothesized that an individual who was intrinsically motivated would be more satisfied with the upward communication in the organization than a person who was extrinsically motivated. This hypothesis was not supported ($F=1.978$, $df=3/61$, $p=.13$).

Peterson (1980,1982) began the development of a conceptual model for upward communication based on Vroom's expectancy-valence model. Specifically, the theory states that people will evaluate a variety of possible behaviors and then choose the particular behavior which they believe will lead to desirable outcomes (Szilagyi & Wallace, 1983.)

Vroom theory (1964) views motivation as a multiplicative function. It is necessary for both expectancies to be high to motivate a person to behave in the desirable direction. This is the relationship depicted in Figure 1.

As we initially visualized the relationship between the expectancy variables and upward communication, we thought that those subjects which reported they were high on both Expectancy I and II would also report a higher level of upward communication. Next, those subjects who reported being low on both expectancy variables would report a lower level of upward communication. Finally, we thought that those people who reported being high on one variable but low on the other variable would report less upward communication than high/high subjects but more than low/low subjects.

METHODS

This research was conducted at the U. S. Air Force Academy. While students usually present a problem in conducting this type of research, cadets represent a unique subject group. While each cadet is a student, (s)he also holds a job in his or her squadron. So while the subjects are students, they are also

subordinates in their squadron and have the opportunity to actively communicate upward if they choose.

Data collection involved the administration of a confidential self-report instrument regarding their expectancies and the amount they communicated upward. The sample consisted of 794 subjects of which 571 usable questionnaires were obtained. Thus, the sample in this paper represents 72% of the total population.

A single self reported item was used to determine the amount each subject believed (s)he communicated upward. The scale was an eight point scale with once a day as the high anchor and never as the low anchor.

In 1984, Peterson and Rohrs (1984) reported the initial development of E I and E II scales to test the above relationships. The Expectancy I scale reflects the person's view that effort on his or her part will lead to the desired performance. It is measured with a nine items with coefficient alpha of .89

The Expectancy II scale reflects a person's perception once (s)he has performed. It measures a person's belief that certain outcomes will occur from the desired performance. Expectancy II was measured with 20 items with a coefficient alpha of .94.

ANALYSIS

The objective of this analysis is to examine the relationship between the expectancies variables and the upward communication outcome variable. A two by three cross-classification table was developed and tested for independence. To develop the cross-classification table, E I and E II were recoded into high and low E I groups and high and low E II groups. These groupings were then combined into a high/high group, a low/low group, and a group which combined one high and one low expectancy factor. The upward communication variable was also recoded into high (anyone who reported communicating upward at least once week) and low (all others) groups. These new variables were then cross tabulated using the Crosstabs program in Statistical Package for the Social Sciences (1983).

RESULTS

Table 1 indicates that Expectancies I and II affect the degree of upward communication. Those subjects who reported they are high on both E I and E II communicated upward more than those who reported they are low on both the E I and E II scales. This statistical test indicates a positive relationship between E I and E II on upward communication. While the relationship between low/low and high/high is as predicted the relationships between low/high's and high/low's does not hold. More will be said about this in the discussion section of the paper.

DISCUSSION

This research shows a relationship between expectancy variables and upward communication. If an individual sees both a positive relationship between effort to performance and performance to outcomes, (s)he will tend to communicate upward more than those who do not see this positive relationship between these expectencies. But this does not mean that people with low E I or low E II scores will not communicate upward, just less often. The implications for supervisors are very important. A supervisor must make sure that his or her subordinate not only receive the outcomes they desire from upward communication, but they must also insure that their subordinates believe that their effort will in fact lead to the desired behavior.

When conceptualizing the relationship, the authors thought that those individuals which fell into the categories of low/high and high/low would display more upward communication than the low/low's, but less than the high/high's. This did not prove to be true. One explanation for these findings is that when questioned on the frequency of upward communication subjects perceive it as a socially desirable behavior and so report communicating upward more than they actually do.

While one must be careful in generalizing these results to the general populace, this research does demonstrate the need to further explore the relationship between expectancy theory and upward communication. Further research is needed in this area which examines the relationship between upward communication, expectancy variables, and other demographic characteristics.

CONCLUSION

This research examined the relationship between upward communication and Vroom's expectancy model of motivation. Results indicated that those people who reported they were high on both E I and E II significantly communicated upward more than those people who reported being low on both E I and E II. This research suggest the need to study expectancy variable as they affect the amount of upward communication demonstrated by a subordinate.

References Available Upon Request.

FIGURE 1

THE RELATIONSHIP BETWEEN EXPECTANCY THEORY AND UPWARD COMMUNICATION

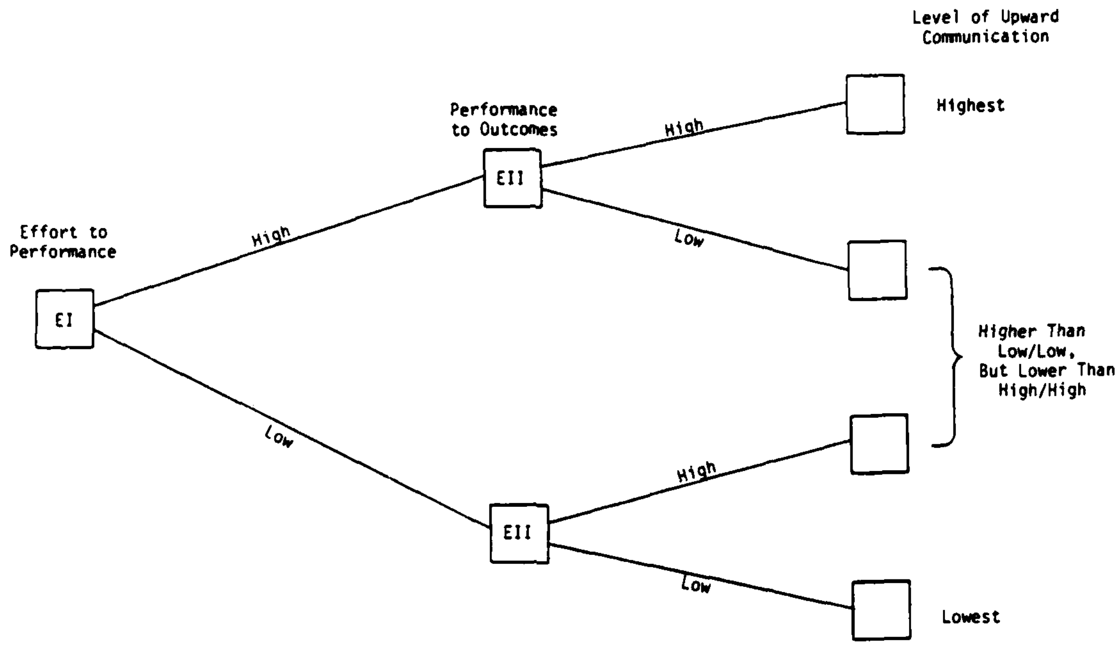


TABLE 1

TWO BY THREE CROSS-CLASSIFICATION TABLE FOR UPWARD COMMUNICATION WITH EXPECTANCY GROUPINGS

	COUNT ROW PCT COL PCT TOT PCT	UPWARD COMMUNICATION LOW	 HIGH	ROW TOTAL
EXPECTANCY GROUPINGS	HIGH/ HIGH	75 42.4 41.2 13.1	102 57.6 26.2 17.9	177 31.0%
	ONE HIGH/ ONE LOW	55 34.6 30.2 9.6	104 65.4 26.7 18.2	159 27.8%
	LOW/ LOW	52 22.1 28.6 9.1	183 77.9 47.0 32.0	235 41.2%
	COLUMN TOTAL	182 31.9%	389 68.1%	571 100.0%

Chi Square = 19.8***
Kendall's Tau C = .187***

* - .05 level of significance
** - .01 level of significance
*** - .001 level of significance

USAFA FOURTH CLASS COMMITTEE -- CHALLENGES AND SUCCESSES

Kenneth H. Fleming, Ildiko E. Andrews, and Robert F. Phelps

United States Air Force Academy
Colorado Springs, Colorado 80840

ABSTRACT

The U. S. Air Force Academy has an extensive and well developed system for monitoring the academic progress of all students. The focal point of this system for each class is a standing committee called the class committee. The class committee is chaired by a permanent professor and consists of instructors who are usually selected from across the spectrum of courses offered in that year. The primary function of the committee is to evaluate the suitability of students for retention at the Academy, and to make a recommendation to the Academy board--the final authority on retention or dismissal. A secondary, but very important function of the committee is to motivate students to improve their academic performance. This is done through interviews, counseling, extra instruction, and a formal study and guidance program. This paper will describe the process of the class committee in some detail, but will focus primarily on the secondary function. A large body of statistical data has been collected on students who have been through the committee process. Summary statistics regression and correlation analysis will be used to analyze the data to ascertain (in so far as this is feasible) whether the committee process improves educational performance. A discussion of the results will be presented together with any recommendations.

INTRODUCTION

The focal point of the academic evaluation system for each of the four cadet classes is a standing committee known as the class committee. Each of the four class committees are chaired by a permanent professor (a tenured academic department head) and consists of instructors and other officers who are selected from across the spectrum of academic, athletic, military, and medical aspects of cadet life. The primary function of the class committee is to evaluate the potential of cadets for retention at the Academy and to make a recommendation to the Academy Board. A secondary function is to motivate or challenge the cadets to improve their academic performance. This task is accomplished through the use of interviews, counseling, referral for such assistance as academic tutoring called "extra instruction", how-to-study programs, or for professional guidance services provided by the Cadet Counseling Center.

METHODOLOGY OF THE COMMITTEE

The four class committees meet at midsemester and end of semester to review academic records of all cadets deficient in academic studies. The freshman class committee meets cadets at one additional grading period, the miniprog, after the first month of fall semester. Those deficient cadets who fall into multiple F, hardlook GPA, or multiple probation categories (as well as selected others) are interviewed. Unless the committee specifically takes action to the contrary, cadets deficient at midsemester will be placed on academic probation. At the end of semester, the class committee will recommend to the Academy Board that an academically deficient cadet be disenrolled from the Academy unless, in the committee's estimation, the cadet's overall performance and potential for successfully completing the academic program justifies a recommendation for retention.

THE CLASS COMMITTEE INTERVIEW

Once a cadet has been identified as deficient in academics, the committee chairman will select those cadets to be interviewed by the full class committee. Normally, cadets with multiple "F" grades, a cumulative GPA of less than 2.00, or on athletic or military probation will be asked to appear before the committee to present information pertinent to their academic performance during the grading period being discussed.

A class committee interview may be subdivided into four phases: Characterization, Inquiry/Advisory, Recommendation, and Counseling. The purpose of the class committee is to determine if the overall performance and the probability of a cadet successfully completing the academic program justifies a recommendation for retention. Therefore, it is incumbent upon the members to collect the facts relating to academic performance and attempt to measure the probability of success.

During the characterization phase, a profile or character sketch of the cadet is composed using facts and observations gleaned from the instructor comment cards, formal records of counseling, medical history, military performance appraisals, and athletic ratings. Quite often, the pattern of evidence from these sources is indicative of a particular problem to be investigated further when the cadet is present during the next phase of the interview process.

Given the opportunity to explain behavior or lack of academic performance, the cadet is the major participant during the inquiry/advisory phase. The function of the committee is to ascertain any factors influencing the cadet's behavior such as family, personal, peer, or academic problems as well as to offer solutions to mitigate these problems.

By focusing on the specific problem just identified, appropriate measures may be suggested to modify and correct those conditions detrimental to successful academic performance. Help is available over a broad spectrum. A cadet may seek extra instruction from an instructor which may be brief lecture reviews, working additional problems, or periodic, structured tutorial sessions. A well developed How-to-Study program has been established which enables cadets to identify and correct many common study deficiencies. Administered under the auspices of the Behaviorial Sciences and Leadership Department, this very successful program enlists the aid of instructors from across the faculty as study counselors. For very serious problems, professional guidance is available from the Cadet Counseling Center and the Cadet Clinic.

Following the class committee interviews, the counseling phase begins as the cadets are interviewed and counseled by committee members on a one-on-one basis. The purpose is to review the class committee action and plan a program to return the cadet to good academic standing.

STATISTICAL EVALUATION

INTRODUCTION

Over the course of the 1984-1985 school year, statistics were compiled on the freshman class committee deliberations. The raw data consists of a number of attributes on all those cadets who were academically deficient at any of the grading periods during the year. Each of these cadets "met the committee" in either a formal interview or in a one on one counseling with a committee member. One exception to this is provided by the cadets who were deficient for the first time at the end of the spring semester. These cadets were not previously deficient and therefore did not receive any formal counseling.

DATA

540 cadets met the committee during the year; of these, 166 were repeaters, so that 374 separate individuals were deficient at some time during the year. Data on each cadet includes his or her academic composite entry score (a weighted average of SAT scores and class standing); whether the cadet was automatically downloaded academically because of a low academic composite; whether the cadet attended a preparatory school; the number and timing of academic deficiencies; semester and cumulative GPAs; and whether the cadet was deficient at the end of the school year.

ANALYSIS

The data are analyzed in two sections: The first covers the "validity" of the selection/identification process, and the second the outcome of the committee process. Certain cadets who have low academic composites are automatically given a lesser academic load (four instead of five courses); other cadets have already had a year in a preparatory school because of low entry scores. Both of these groups can be thought of as pre-identified students who may have academic problems (see Table 1). The committee process itself begins the identification of cadets during the academic year.

	NUMBER/TOTAL	
DOWNLOADED	185/1522	8.2%
PREP SCHOOL	196/1522	7.8%

Table 1. Students who are pre-identified for academic problems. Twenty-four students were both downloaded and prep school; these are all included in the prep school number to avoid double counting.

PRE-IDENTIFICATION

The percentage of downloaded and prep school cadets who are deficient shows the performance of these groups. Table 2 gives the percent who were deficient at the grading periods:

	MID FALL	END FALL	MID SPRING	END SPRING	END YEAR
DOWNLOADED	13.5	4.7	23.0	18.0	21.3
PREP	25.4	14.3	17.3	27.2	21.3

Table 2. Percent of deficient students who are downloaded or prep.

Based on the raw percentages in Table 2, it is clear that the selection process has "identified" those cadets who are likely to have academic problems. Prep school and underloaded cadets account for forty three percent of the deficiencies at the end of the year and make up only 16% of the student population. The same disparities show up across the table. The percentages also show that the underloaded program "helps" during the first semester; that is, these students are less likely to show up as deficient at the end of the fall semester. Prep school has a small positive effect in the first semester; interestingly, both Prep school and download have the same percentage of deficiencies at the end of the year. While instructive, percentages are only a first look at these students--a more interesting question is how much, if any, the underload and prep school help these cadets. To answer this question refer to Tables 3 and 4. In Table 3, we have run an ordinary least squares (OLS) regression of those students who are deficient at the end of the year against the explanatory variables: academic composite, underload, prep school, and whether or not the individual met the committee during the year. The OLS procedure is a rough attempt to separate out the

effects of the different independent variables. A full discussion of the technique (and its limitations) is beyond the scope of the paper. See Hanushek for a complete discussion. The coefficients on the variables can be interpreted as the increment (+ or -) to the probability of being deficient at the end of the year. Consider Table 3 and take an individual who does not meet the committee at all during the year (committee effects will be discussed in the next section).

DEPENDENT VARIABLE	INDEPENDENT VARIABLES				COMMITTEE MEETINGS		
	CONSTANT	DWNLD	ACAD COMP	PREP	MID FALL	END FALL	MID SPRING
DEFICIENT AT END OF YEAR	.886	+.138	-.000159	.035	-.137	.201	.277
LEVEL OF SIGNIFICANCE		.95	.92	NOT	.97	.97	.99

Table 3. Ordinary least squares (OLS) regression relating end of year deficiency to explanatory variables.

If we consider no other effects, this individual has a .88 probability of being deficient at the end of the year (the constant term). We note, however, that academic composite has a substantial negative effect on the probability of being deficient. If we award our hypothetical student the average academic composite of this sample (2900) then his or her probability of being deficient is (.886 -.000159 (2900)) or .425. Every extra one hundred points of academic composite decreases the probability of deficiency by 1.6%; for example, an individual with an academic composite of 3200 is 8% less likely to be deficient than one with an academic composite of 2700. Consider the effects of downloading and prep school respectively. Downloading adds 13.8% to the probability of being deficient and is statistically significant, whereas prep school adds only 3.5% to the probability and is not significant. Since deficiency can come about in a number of ways, (for example, one or more Fs, or low GPA), we might ask if downloading does anything toward improving GPA. To answer this question refer to Table 4 which gives the results of a least squares regression of grade point average against the same independent variables as Table 3.

DEPENDENT VARIABLE	INDEPENDENT VARIABLES				COMMITTEE MEETINGS		
	CONSTANT	DWNLD	ACAD COMP	PREP	MID FALL	END FALL	MID SPRING
GPA AT END OF YEAR	1.219	+.146	-.00030	.01	-.19	-.33	-.19
LEVEL OF SIGNIFICANCE		.97	.99	NOT	.99	.99	.99

Table 4. OLS regression relating end of year GPA to explanatory variables.

If we consider the effect of downloading on GPA, it raises the GPA of these students (relative to all other deficient students) by .146. Thus downloading helps GPA, but not enough to keep the downloaded individual out of the deficient category. Prep school has no pronounced effect one way or the other.

The data clearly indicate that the early measures of academic success are valid in so far as they identify students who may have difficulty. In every equation, academic composite is a large and statistically significant factor. Underloaded students are also a high risk group; however, the underload does help grade point average. Preparatory school appears ambiguous since prep school students make up a large percentage of the deficiencies but the

coefficient of prep school is small and not statistically significant in the causal regressions.

COMMITTEE RESULTS

On a more specific level, we would like to know if the committee system has any motivational effect. We have already seen that the entry flags are valid and that the remedial measures (downloading and prep school) have a positive (but small) remedial effect. In the case of the prep school, the positive effect is not measurable but can be deduced from the fact that prep school has no causal effect on end of year deficiency. Turning again to Table 3, we look at the coefficients for students who meet the full class committee. The mid-fall committee meeting produces a 13.7% decrease in probability of deficiency at the end of the year. Endfall and midspring have positive and significant coefficients. This suggests that the students who meet the committee in the mid-fall benefit from the procedure--at least as measured by their end of year performance!

FIRST TIME DEFICIENT

There is one other identifiable group that helps explain the motivational aspect of the learning problem. In the spring semester we have thirty six students who were deficient for the first time during the year. This group has not previously been identified or counseled, and Table 5 shows an OLS regression of this group against our earlier predictors.

DEPENDENT VARIABLE	INDEPENDENT VARIABLES		
	DWNLD	PREP	ACOMP
DEF FIRST TIME AT END OF SPRING	.055	.014	.00000668
SIGNIFICANCE	NOT	NOT	NOT

Table 5. First time deficient students.

These results are interesting when we compare them to the results of Table 3. The committee process appears to "help" (if only slightly) those students who are deficient for the first time in the fall. On the other hand, the group from table 5 has received <u>no</u> counseling and they make up ten percent of the total deficiencies in the spring. Since none of our prior predictory are significant we conclude that this group "slipped" into deficiency. It is at least possible that if this group had been counseled (as the fall group was) then they might have avoided deficiency in the spring.

SUMMARY

The data clearly show that the pre-identification process for academically weak students is valid--remedial measures help, but do not have a very large effect. There is evidence to indicate that the committee process helps during the fall semester, and weak indirect evidence at the end of the spring semester, but the data are generally inconclusive on the motivational results of the committee.

REFERENCES

Hanushek, Eric A. and Jackson, John E. 1977 Statistical Methods for Social Scientists Academic Press Inc. New York NY.

Pindyck, Roberts and Rubinfeld Daniel L. 1981 Econometric Models and Economic Forecasts 2nd Ed McGraw-Hill Book Co. New York NY.

Toward An Electronic Bibliography of Military Leadership

David D. Van Fleet and Tim O. Peterson
Texas A&M University

Abstract

A new research monograph contains an extensive bibliography of nearly 4,000 references on military leadership. Like most bibliographies, its utility is limited since it is not coded by content. Further, even if it were coded, searching for sets of particular references (all of those using Army samples for company grade officers, for instance) would be very time consuming. This paper suggests converting that bibliogrpahy into a database and making it available to users through an electronic bulletin board. Advantages, applications, and some limited examples are noted.

Introduction

In their forthcoming monograph on military leadership, Van Fleet and Yukl (1986) include a bibliography with nearly 4,000 references. That bibliography was developed to be the single, most complete bibliography on military leadership available. Researchers and practitioners alike should find the bibliogrpahy valuable as a guide to sources of information as well as to ideas and examples.

Development

To provide integration and/or cross-fertilization, efforts were made to include all branches of the military by gathering references from military libraries. Those libraries were asked to identify everything in their card catalogs under the headings, "leadership" and "military leadership." In addition to references identified in this way, other published bibliographies and technical reports provided material used in the bibliography. Libraries which participated in the development of this bibliography include the following:

Academy Library
USAF Academy
USAF Academy, Colorado

Air University Library
Maxwell Air Force Base, Alabama

Schuyler Otis Bland Memorial Library
U.S. Merchant Marine Academy
Kings Point, New York

James Carson Breckinridge Library
U.M. Marine Corps
Quantico, Virginia

Carlisle Barracks Library
U.S. Army War College
Carlisle Barracks, Pennsylvania

Combined Arms Research Library
Ft. Leavenworth, Kansas

Eglin Base Library
Eglin Air Force Base
Eglin AFB, Florida

Headquarters Library
Air Systems Command
Andrews AFB
Washington, D.C.

National Defense University Library
Washington, D.C.

Naval War College Library
Newport, Rhode Island

Nimitz Library
U.S. Naval Academy
Annapolis, Maryland

Technical Library
U.S. Naval Air Systems Command
Washington, D.C.

U.S. Coast Guard Academy Library
U.S. Coast Guard Academy
New London, Connecticut

U.S. Military Academy Library
U.S. Military Academy
West Point, New York

Wheeler Base Library
Wheeler AFB, Hawaii

Coding

Most of the bibliographic references are about military leadership, but some are business/industry references. Some are historical and/or biographical, some are foreign, and some cannot readily be classified. Like all bibliographies, however, finding references to subjects in which you are interested is difficult since they are arranged alphabetically and not by subject.

What is needed is to have the references coded in a way which would make them more useful to users. A tentative coding scheme has been developed. Using this scheme, a user could easily tell a lot about the contents of a reference. By identify those with the same code for a particular topic, users could locate all references of use to them. The coding scheme uses a seven category system as follows:

Subject Scope
- CC combat/emergency command leadership
- NC non-combat/non-emergency command leadership
- CA combat/emergency administration leadership
- NA non-combat/non-emergency administration leadership
- LS leader selection
- LT leader training and education
- MP military profession
- XC not subject scope specific; unclassifiable

Theory Basis (alphabetical)
- EX exchange
- IE interaction/expectation
- HU humanistic
- PT personalty trait or "great man"
- SC situational/contingency
- XT not identifiable; unclassifiable

Model Basis (alphabetical)
- Grid Managerial Grid or Blake and Mouton
- LPC Fiedler or LPC
- Mich Michigan or Employee-centered and Production-centered
- Ohio Ohio State or Initiating Structure and Consideration
- Path Path-Goal
- Sys4 Systems 1, 2, 3, 4
- VY Vroom-Yetton
- Yukl Yukl's taxonomy or framework
- XM not model specific; unclassifiable

Content of Material (alphabetical)

A/C	analysis/critique
B/A	biographical/autobiographical
H-T	guide or "how to"
MN	service manual
PO	personal opinion
RS	research based, empirical report
TD	theoretical development
XC	not content specific; unclassifiable

Leadership Level

UO	officer corps unrestricted
GO	general grade (admiral)
FO	field grade (Col, Lt Col, Mjr; Navy-Cpt, Cdr, Lt Cdr)
CO	company grade (Cpt, 1 Lt, 2 Lt; Navy-Lt, Lt JG, Ens)
UNCO	NCO corps unrestricted
SNCO	senior grade NCO
JNCO	junior grade NCO
XL	not level specific; unclassifiable

Service (alphabetical)

AF	Air Force
AR	Army
CG	Coast Guard
MR	Marines
MM	Merchant Marine
NV	Navy
PM	para-military (police, fire, etc.)
XS	not service specific; unclassifiable

Group Size

SG	small group	(less than or equal to 15)
MG	medium group	(between 15 and 200)
IG	intermediate group	(between 200 and 1,000)
LG	large group	(between 1,000 and 5,000)
VG	very large group	(between 5,000 and 20,000)
OG	organizational size group (over 20,000)	
XS	not size specific; unclassifiable	

While adding this coding information to each reference would be a considerable improvement, it would still be a laborious task for users to sort through the 4,000 entries. It would not be, however, if the bibliography were available in an electronic format. In an electronic format, users could search rapidly for any one or even combinations of coded references. This would not only greatly speed up searches but also enable users to extend their work since the effort to locate references would be reduced.

Electronic Bulletin Board

The most useful way in which to make this bibliography available in an electronic format would be as a bulletin board system (BBS). In this format, users anywhere in the world could, via a modem, telephone the BBS, search by one or more codes, and see the references in which they are interested. Such a system could also enable them to obtain printed copies of the materials and to inform the operators of the system about additional materials which could be added to keep the bibliography up-to-date and complete.

That system is currently under development by the authors using a database system to further extend its usability. The authors hope to obtain funding to offer the system on a trial basis by August 1986. Anyone interested in making suggestions about the suggested coding scheme or the electronic bibliography should contact the first author.

Content Analysis (A through E)

As an example of one very limited use of such an electronic system and to provide some limited information about the bibliography, a content analysis has been conducted of the A through E portion of the bibliography. That analysis reveals that journal articles are now far more frequent than are books.

Date	Books	Journal Articles
pre 1900	17	0
1900-1929	19	8
1930-1939	10	5
1940-1949	24	24
1950-1959	46	71
1960-1969	96	68
1970 on	115	226

The five most frequently referenced journals are: Journal of Applied Psychology (186 references), Military Review (31 references), U.S. Naval Institute Proceedings (15), Army (12), and Personnel Psychology (11). The five most frequent publishers for books are: U.S. Government Printing Office (22); Harper & Row (12); Little, Brown (8); McGraw-Hill (7), and Stackpole/Military Service Publishing (7).

The five most frequently referenced authors for journal articles are: Bass (9), Beaumont (7), Csoka (7), Fiedler (6), and Argyris (5). Ignoring service authorship, the five most frequent book authors are: Argyris (6), Eisenhower (6), Churchill (5), Corbett (5), and Bass (5).

The frequency of word use in titles is also interesting. The 15 most frequently used words in journal titles and the 10 in book titles (ignoring prepositions, conjunctions, and articles) are shown here.

word	frequency in articles	frequency in books
lead/leader (+ suffixes)	162	73
manage/manager (+ suffixes)	42	25
military	42	34
effect/effective (+ suffixes)	40	12
group(s)	33	9
relation (+ suffixes)	30	0
behavior (+ suffixes)	28	8
command (+ suffixes)	27	13
train (+ suffixes)	25	9
officer(s)	23	21
develop (+ suffixes)	18	5
organization (+ suffixes)	18	19
combat or war (+ suffixes)	17	24
army	15	19
attitude (+ suffixes)	14	2

Word frequency information such as this could be categorized and compared with the frequency of those words in English language generally for a true content analysis (see Krippendort, 1980; Kucera and Francis, 1967; Weber, 1985; Wood, 1980). This form of analysis could be suggestive of how researchers on military leadership conceptualize that construct. If the sample were large enough, this form of analysis could be done by service to compare and/or contrast conceptualizations.

Summary

This overview suggests what could be done with an electronic bibliography on military leadership. Initial work has been completed on such a system and that early version will be demonstrated. This demonstration will provide the audience time for questions, comments, and suggestions which could be used to assure that the system, when fully developed and implemented, will be of value to and used by researchers in the Department of Defense as well as other organizations.

References

Krippendort, K. Content Analysis: An Introduction to its Methodology. Beverly Hills, CA: Sage Publications, 1980.

Kucera, H., & Francis, W.N. Computerized Dictionary of Present-Day American English. Providence, RI: Brown University Press.

Van Fleet, D. D., & Yukl, G.A. Military Leadership: An Organizational Behavior Perspective. Greenwich, CT: JAI Press, 1986.

Weber, R.P. Basic Content Analysis. Beverly Hills, CA: Sage Publicationsk, 1985.

Wood, M. "Alternatives and options in computer content analysis." Social Science Research, 1980, 9, 273-286.

Family and Work in the Air Force II:
The Air Force Family Survey

Mickey R. Dansby
Leadership and Management Development Center

Philip Lewis and Dorothy Winther
Auburn University

Abstract

Recently, the Air Force Leadership and Management Development Center (LMDC) revised its USAF Spouse Survey (AFSS) and renamed it the USAF Family Survery (AFFS). The present paper describes the factor analyses of the AFFS and the resultant factor structure. Regression models are developed predicting Air Force members' career intent and job attitudes using AFFS factors and variables as predictors. Results using the AFFS are compared to previously reported (Dansby, 1984) regression results using the AFSS. Although predictive ability for career intent and job attitudes was similar for the two surveys, the AFFS contributes for unique and statistically significant ($p < .05$) predictor factors: stressfulness for the family of Air Force lifestyle, family life cycle, spouse's marital satisfaction and spouse's social isolation. These factors are of considerable significance in the family stress and coping literature.

In 1984, in an attempt to provide better information about the interrelationship of family and work in the Air Force, the Leadership and Management Development Center (LMDC) initiated a revision of its family research instrument, the U. S. Air Force Spouse Survey (AFSS; Dansby, 1984; Flannery, 1985). The resulting instrument, renamed the Air Force Family Survey (AFFS), represents a systematic attempt to improve the previous instrument both by improving its psychometric qualities and by adding items that sample areas of theoretical and empirical relevance that were not included in the previous instrument. Because of these improvements, it was expected that the new AFFS would provide a more complete understanding of the family-work interface than had the previous AFSS. The present paper is a brief summary of the factor structure of the new instrument and its performance relative to the previous instrument in predicting Air Force members' career intent, job satisfaction, and perceived work group effectiveness.

Method

The factor structure of the AFFS was based on the matched responses of 1170 Air Force members and their spouses for a total of 2340 people. Fourteen percent of the Air Force members were officers and 84 percent were enlisted personnel. For each pair, the Air Force member had completed the Organizational Assessment Package (OAP), a questionnaire used by the LMDC to measure work-related attitudes. The member's spouse subsequently completed the AFFS. Return rate for the AFFS was 35%. Individuals completing the OAP were part of a census of the military personnel in large organizations at three different Air Force bases. Since the bases were not randomly selected, the sample of matched pairs is an "opportunity" sample that cannot necessarily be generalized to all domestically based Air Force families. Nonetheless, it is a relatively large sample and it has the advantage over much previous research of including the responses of both spouses in these Air Force families. For this reason these data permit a direct examination of relationships between the attitudes of military personnel and their spouses.

Factor Structure of the AFFS

The Air Force Family Survey (AFFS) is a 140 item questionnaire that is divided into three main sections. The first 61 items assess the spouse's attitudes a about variety of topics ranging from involvement in the Air Force lifestyle to marital satisfaction and the spouse's gender role orientation. In the middle section are four groups of items concerning 1) frequency of the family's use of various Air Force services, 2) satisfaction with those services, 3) factors perceived by the spouse to be affecting the Air Force member's career intent, and 4) the perceived impact of various factors connected with Air Force life on the family. The final section of the AFFS consists of 23 items that assess demographic characteristics of the spouse or the family. Items from two sections of the AFFS were subjected to factor analyses: the family attitude items (items 1-61) and

responses to items having to do with the Air Force member's career intent and the impact of certain Air Force factors on the family.

For the 61 attitudinal items a factor structure was arrived at on the basis of three considerations: 1) initial and confirmatory factor analyses (Long, 1983), 2) theoretical considerations (Dansby, 1984), and 3) a requirement that each item be included in one and only one factor. The initial method was a principal component analysis followed by a varimax rotation. Using Kaiser's criterion (eigenvalue greater than one), 17 factors were extracted of which 14 were easily interpretable. All 61 items were allocated to one of the 14 factors; and, with the exception of three items, all had factor loadings of at least .27. Following this procedure, a confirmatory factor anaylsis was conducted using a maximum likelihood factoring method where 14 factors were specified a priori. This was followed by a varimax rotation to obtain a final factor structure. Twelve of the 14 factors identified in the exploratory analysis were confirmed in the second analysis. The confirmatory factor analysis divided two of the initial fourteen factors into two factors, combined two of the original 14 factors into a single factor, and failed to identify one of the inital factors. In all, 53 of the 61 items were allocated to the 12 factors identified by both analyses. Because there were good theoretical reasons for retaining the factor not confirmed in the second analysis (Olson's cohesion factor; Olson, Sprenkle & Russell, 1979) and for retaining as a single factor the factors which were split on the second analysis, the original 14 factors were accepted as the variable structure for use in subsequent analyses. Factor scores were obtained via simple linear combination of item responses divided by the number of items in the factor. Table 1 lists the 14 factors and their internal consistencies (Cronbach's alpha).

Table 1
AFFS Attitudinal Factors

Factor	AFFS items*	Name	# items	Alpha
1	13,14,16,22,27,29,36	Air Force Member's Job Stress	7	.72
2	12,19,21,33,34,39	Stress of AF Life for the Family	6	.60
3	28,37,57,58,-59,60,61	Marital Satisfaction	7	.75
4	2,5,7,18,26,32	Positive View of the Air Force	6	.80
5	8,10,23,25,41	Sensitivity of AF to Family Needs	5	.75
6	1,3,4,6,9,11,-47	Commitment to AF Lifestyle	7	.74
7	15,20,30,38,-40,42	Perceived Job Satisfaction	6	.72
8	24,43,44	Spouse's Career Orientation	3	.36
9	45,46	Gender Role Orientation of Spouse	2	.57
10	52,55	Spouse Independence	2	.46
11**	48,-49	Family Disengagement	2	.59
12	31,35	Member's Career Intent	2	.85
13	17,-51,53	Social Isolation	3	.55
14	50,-54,56	Help Seeking Attitudes	3	.40

*A negative sign indicates that the item loaded negatively on the factor.
**This factor did not emerge on the confirmatory factor analysis.

A second analysis was conducted to determine the factor structure of AFFS items 99-117, the items concerning the Air Force member's career intent and the spouse's view of the effect of a number of Air Force factors on family life. A principal component analysis followed by a varimax rotation yielded four "clean" and easily interpretable factors. Factor loadings of the individual items ranged from .41 to .78, with all items loading on at least one factor. Those factors were perceived favorableness of basic job benefits (items 101, 102, 104, 105, 106, & 108), perceived impact of Air Force moves, exercises, and temporary duty assignments (items 107, 109, 111, 112, & 115), perceived job satisfaction (items 99, 100, 103, & 110), and satisfaction with the current Air Force duty location (items 113, 114, 116, & 117). As with the 14 family attitude factors, factor scores were obtained via simple linear combination of item responses divided by the number of items in the factor.

Predicting Air Force Member's Career Intent and Job Attitudes

Using the earlier AFSS, Dansby and Hightower (1984) were able to predict Air Force members' career intent, job related satisfaction, and perceived work group effectiveness with varying degrees of success (multiple Rs ranged from .362 to .745). They concluded that their findings lent support to the proposition that family and spouse attitudes and perceptions have an important impact on Air Force members' work attitudes. Dansby and Hightower also predicted that additional associations between family and work would be revealed by a family assessment instrument that sampled additional theoretically important variables. The AFFS, with its greater array of factors pertaining to family structure and dynamics, should permit a test of that prediction.

To provide a comparison of the Spouse Survey (AFSS) and the new Family Survey (AFFS) the multiple regression analyses conducted by Dansby and Hightower were repeated using the AFFS and the same or similar predictor variables as those used previously. As with the Dansby and Hightower regression analysis, the spouse's perception of the Air Force member's career intent was not included as a predictor variable since it was expected to be highly correlated with two of the three criterion variables (member's career intent and job related satisfaction) and does not directly assess family or spouse attitudes. A "full" regression model was constructed first, which included all the predictor variables specified by Dansby and Hightower (1984). This was followed by identification of a "restricted" regression model which included only those predictor variables in a forward stepwise inclusion procedure where the criterion for inclusion was $p < .05$.

Regression Results

The regression models for the AFFS were developed on the matched responses of 811 of the 1170 Air Force members and their spouses in the present sample. Excluded were Air Force members within one year of retirement or who were married to another Air Force member, since it was expected that the dynamics of their family-work relationships would be different from those who were included. Tables 2, 3, and 4 present a comparison of the full and restricted regression models for each of the three criterion variables, where predictions from the Spouse Survey are compared with predictions from the Family Survey.

Table 2
Full and Restricted Regression Model Results for the AFSS and AFFS in Predicting Career Intent

	FULL MODEL				RESTRICTED MODEL			
Survey	df	F	R	R^2	df	F	R	R^2
AFSS								
Regression	36	34.6	.533	.284	10	120.4	.525	.276
Residual	3137				3163			
AFFS								
Regression	31	15.6	.618	.383	10	47.1	.609	.370
Residual	780				801			

Table 3
Full and Restricted Regression Model Results for the AFSS and AFFS in Predicting Job Related Satisfaction

	FULL MODEL				RESTRICTED MODEL			
Survey	df	F	R	R^2	df	F	R	R^2
AFSS								
Regression	36	108.4	.745	.554	8	540.9	.738	.546
Residual	3137				3165			
AFFS								
Regression	31	16.9	.634	.401	4	123.9	.617	.381
Residual	780				807			

Table 4
Full and Restricted Regression Model Results For the
AFSS and AFFS in Predicting Perceived Work Group Effectiveness

	FULL MODEL				RESTRICTED MODEL			
Survey	df	F	R	R^2	df	F	R	R^2
AFSS								
Regression	36	13.1	.362	.131	7	64.1	.352	.124
Residual	3137				3166			
AFFS								
Regression	31	3.7	.358	.128	4	25.6	.336	.113
Residual	780				807			

Although comparisons of results with the two different Air Force surveys of family and spouse variables should be treated with caution (given the fact that the AFSS and AFFS were collected on different individuals at different Air Force bases using samples of diferent sizes), some aspects of the comparisons seem particularly noteworthy. First, the multiple regression models for the three criterion variables show the same rank ordering in terms of the size of the multiple Rs produced. Second, these multiple Rs are of roughly equivalent magnitudes for the two different surveys. In both studies, the questionnaire responses of the spouse were most predictive of the Air Force member's job related satisfaction (Rs of .745 and .634) and least predictive of perceived work group effectiveness (Rs of .362 and .358). While it is not possible to provide a statistically meaningful comparison of the multiple Rs obtained, it is possible to make a comparison of the regression models developed using the two different surveys to determine which is more effective in predicting the criterion. This is done by constructing an F ratio using the mean square residual from the two models (Sympson, 1980). When this is done, the Spouse Survey is found to be more effective in predicting Job Related Satisfaction than is the newer Family Survey. The F ratios for the other two criterion variables were quite low (1.11 and 1.26), suggesting that the two surveys are about equally effective in predicting career intent and perceived work group effectiveness.

Following the procedure used by Dansby and Hightower (1984), regression models excluding the one predictor variable not derived from the spouses' survey responses (the Air Force member's perception of the job support received from the family) were also computed for the AFFS. As was true in the previous study using the AFSS, substantial reductions in predictive ability occurred for job related satisfaction (the multiple R dropped from .634 to .427) and perceived work group effectiveness (the multiple R dropped from .358 to .257). The reduction was very slight for career intent (from .618 to .602).

Finally, several factors and items were added to the list of predictor variables to see if they increased the ability of the AFFS to predict the three criterion variables. These variables were AFFS factors 4, 5, and the factor pertaining to satisfaction with the current duty location (see Table 1), and items assessing living location (on or off base), the spouse's educational level, student status, and degree of involvement in volunteer work. The addition of these variables to the prediction models did not change the multiple R for career intent, but it did raise the mulitple Rs for job related satisfaction and perceived work group effectiveness. In the restricted model, with the Air Force member's perception of family job support removed, spouse's education level became a significant ($p < .05$) predictor of perceived work group effectiveness and living location became a significant predictor for job related satisfaction. However, the simple correlations of these predictors with the criterion variables were quite low (.08 and .05) and the increases in the multiple Rs small (.03 and .01).

One of the purposes for developing the AFFS was to provide a measure of Air Force spouses and families that included family variables that have been identified in the empirical and theoretical literature as critical to an understanding of the family-work interface in the Air Force. If this purpose has been achieved, then we would expect differences in the nature of the predictors found using the newly developed AFFS as compared to the predictors found with the older AFSS. This was indeed the case. Looking at the Air Force member's career intent, we find that there were ten significant predictors for both the AFSS and the AFFS. Six of the ten were identical for the two surveys. Three of the four unique predictors from the Spouse Survey pertain to concrete aspects of Air Force life: Satisfaction with recreational facilities, perceived job benefits, and TDY frequency. In contrast, the four unique predictors from the new Family Survey were variables of considerable interest in the family dynamics literature: 1) the perceived

stressfulness of the Air Force lifestyle for the family (Factor 2); 2) the absence of school age children at home, a family life cycle variable (McGoldrick & Carter, 1982); 3) the spouse's level of marital satisfaction (Factor 3); and 4) the spouse's level of social isolation (Factor 13). All are variables considered to be important in the literature on family stress and coping (McCubbin, Cauble, & Patterson, 1982).

Discussion

The present findings lend further support to the growing empirical literature on the relationship between work life and family life (see recent reviews by Beeson, 1985; Greenhaus & Beutell, 1985; Hunter, 1982; and Orthner, 1980). Although the nature and directionality of the relationships between family and work have not been fully determined, the development of the Air Force Family Survey appears to have considerable promise for investigating the family side of the equation. The Family Survey has a factor structure that includes a number of variables of considerable interest in the family development literature (Walsh, 1982), including: the perceived stressfulness of Air Force life for the family; marital satisfaction; the spouse's career orientation, gender role orientation, independence and social isolation; and the family's level of cohesion and help-seeking attitudes. None of these variables is included on the Spouse Survey. The AFFS also permits one to determine a family's life cycle stage. Results in the current study of significant unique AFFS predictors of the Air Force member's career intent show that several of these family variables do indeed relate significantly to important work experiences. Additional research is underway using the AFFS and OAP to try to clarify further the role of family and spouse issues in the work experiences and job commitments of Air Force members.

References

Beeson, G. W., Jr. (1985). Influence on the identification of wives with the Air Force organization: An examination of the two-person career pattern within the military. Unpublished doctoral dissertation, University of North Carolina, Greensboro, NC.

Dansby, M. R. (1984). A proposal for the revision of the U. S. Air Force Spouse Survey (Report No. 84-0645). Maxwell AFB, AL: Air Command and Staff College.

Dansby, M. R., & Hightower, J. M. (1984). Family and work in the Air Force. In G. E. Lee & T. E. Ulrich (Eds.), Proceedings, Psychology in the Department of Defense, Ninth Symposium (Report No. USAFA-TR-84-2) (pp. 455 - 459). Colorado Springs, CO: United States Air Force Academy.

Flannery, P. A. (1985). U. S. Air Force Family Survey revision (Report No. 85-0825). Maxwell AFB, AL: Air Command and Staff College.

Greenhaus, J. H., & Beutell, N. J. (1985). Sources of conflict between work and family roles. Academy of Management Review, 10, 76-88.

Hunter, E. J. (1982). Families under the flag: A review of military family literature. New York: Praeger.

Long, J. S. (1983). Confirmatory factor analysis: A preface to LISREL. Beverly Hills, CA: Sage.

McCubbin, H. I., Cauble, A. E., & Patterson, J. M. (1982). Family stress, coping, and social support. Springfield, IL: Charles C. Thomas.

McGoldrick, M., & Carter, E. A. (1982). The family life cycle. In F. Walsh (Ed.), Normal family processes (pp. 167 - 195). New York: Guilford.

Olson, D. H., Sprenkle, D., & Russell, C. (1979). Circumplex model of marital and family systems I: Cohesion and adaptability dimensions, family types and clinical application. Family Process, 18, 3-28.

Orthuer, D. K. (1980). Families in blue: A study of married and single-parent families in the U. S. Air Force (Contract No. F33600-79-C-0423). Greensboro, NC: Family Research and Analysis, Inc.

Sympson, J. B. (1980). Testing differences between multiple correlations (Research Report 79-20). Princeton, NJ: Educational Testing Service.

Walsh, F. (1982). Normal family processes. New York: Guilford.

ABSTRACT

PSYCHOLOGICAL DISTRESS AMONG AIR FORCE WIVES

Marianne M. Pierce, College of the Southwest, Hobbs, NM
M. Lou Luchsinger, University of Baltimore, Baltimore, MD

A psychological study of 332 military wives was conducted in a small southwestern Air Force community. The areas of interest were: (1) the psychological distress of wives in the community, and (2) needs which may be related to psychological symptoms. The study further assessed the relationship of level of education, number of close friends, and rank of husband with psychological distress and needs.

Wives of the Air Force community reported significantly more psychological distress than the BSI's non-patient norm group. In addition, wives of enlisted personnel reported significantly more psychological distress than wives of officers.

Recommendations called for a similar study to be conducted at other bases, further research on coping skills and deficits among military wives, the development of a mental health education program, development of programs in the area of intellectual, social and vocational competencies, and a more comprehensive military community mental health delivery system.

PSYCHOLOGICAL DISTRESS AMONG AIR FORCE WIVES

Marianne M. Pierce, College of the Southwest, Hobbs, NM
M. Lou Luchsinger, University of Baltimore, Baltimore, MD

INTRODUCTION

During the last half of the twentieth century, the institutional military model, one which expects self-sacrifice, dedication to the military as a way of life, and encompasses the entire family, has undergone several major changes. The number of married servicemen, military wives, and dependents increased to more than half of the total force (Bowen, et al, 1983); roles in the civilian society were changing, particularly for women and, in course, affected the traditional roles of military wives. In keeping with these and other trends, an "occupational model" of the military emerged (Moskos, 1970), typified by greater self-interest on the part of military members and family members. Individuals and families who endorsed this model were likely to have significant needs unmet by the military institution. To the extent this model is adopted by military personnel, the problems and needs of spouses become issues of importance in the maintenance of a military force. Thus, the study of military wives' lifestyle problems and their resulting stresses may contribute significantly to the quality of working life of the military man.

Lifestyle Problems of Military Wives

Military psychologists have discussed problems experienced by the wife as a result of her husband's military life affiliation. Often unprepared for the demands of military life, the wife faces several dilemmas. She must attempt to be independent (self-sufficient in needs for companionship) yet dependent (able to leave friendships in order to remain with her husband). She suffers from role ambiguity, role strain, and role conflict due to differing societal and military organizational messages. She may be affected by dissatisfaction with military social life and protocol. Wives experience little sense of community belonging and often feel alienated from their coharts. They have few social contacts outside the military community (Sobie, 1981).

Military wives are not immune to marital and other family problems. For example, frequent geographic moves can be damaging to family structure. Family separations are too long and too frequent, and many wives have difficulty coping with these separations.

Many wives perceive their husband's career commitments as excessively demanding with husbands having a stronger loyalty to the military than to their marriages. Parent-child interactions are reported to be "inconsistent" with problems of care and discipline. Self-identity is at risk with the military wife since there is much role ambiguity and conflict.

Psychological Problems and Syndromes of Military Wives

The literature regarding the mental health of military wives is less extensive than that on other problem or need areas, although the two areas often seem intertwined. Investigations normally focus upon specific problem areas. Wives have been found to suffer from guilt and depression. Wives report severe loneliness and feel abandoned by their husbands. They cannot tolerate being alone. MacIntosh (1968) found wives to suffer from personality disorders and from a variety of psychotic and neurotic disorders, while Jacobson (1980) found a great deal of anxiety disorder among military wives.

Belt and Sweeney (1973) found the lowest levels of psychological adjustment to exist among first-term enlisted wives. In some ways, officers' wives seemed better adjusted than did enlisted men's wives. Women with stronger dependency needs were more likely to show more psychological problems in the military community. According to authors such as MacIntosh (1968), wives who are younger, have less education, and who have less status by virtue of their husband's rank are likely to have lower levels of psychological adjustment.

Problems in Assessment of Psychological Distress

Due to conflicting results of prior research and more recent changes in the military and in women's roles, it is not clear which groups of military wives might be at risk for psychological distress. (Distress is defined in this study as damaging or unpleasant stress) (Selyes, 1974). However, in general military wives having fewer social interactions and lower levels of social support are expected to display more symptoms (Miller and Ingram, 1976).

The Dohrenwends (1965) cite several methodological factors which contribute to wide variance in the study of psychological distress. All have to do with what constitutes a "case." Three dominant methods to distinguish a case can be specified: (1) any person in psychiatric treatment; (2) any person diagnosed by a clinician as having a psychological disorder; (3) any person diagnosed from a standardized questionnaire as having a psychological disorder. The third method, using a standardized questionnaire, is the most preferred since the first two are often unreliable and inadequate.

Most research on the military family is methodologically flawed. It is often based only on cases seen in military medical or psychological practices.

Samples constitute another flaw in research on military wives. Most research is characterized by nonrandom or very small samples.

PURPOSE OF THE STUDY

The present study investigated the psychological states of wives in a military community and the needs which may be related to psychological distress.

METHOD

Subjects

Subjects were 332 wives of United States Air Force personnel stationed at and residing in the area of an Air Force base in the Southwestern United States in a rural area near a medium size city. The respondents represented a response rate of 71% of the original sample which was a stratified random sample of Air Force wives at the base. The modal respondent was a 24-year-old white high school graduate who had been a military dependent about four years. She was an enlisted man's wife, with one to two children, and lived off the base. Her husband planned to make a career of the military. She was as likely to be employed as not, but if she was employed, she was likely to be working full-time. The typical respondent did not belong to her military wives club and described herself as having one to two close friends at her present location.

Instruments

Two instruments were used to investigate needs and psychological distress: the Needs Assessment Scale developed by the senior author, and Derogatis' (1975) Brief Symptom Inventory (BSI). The self-report scale contains 61 statements about military life which the respondent can rate on a five-point scale of distress (0-4). The BSI consists of 53 items and has nine symptom dimensions.

Procedure

Subjects were mailed a questionnaire packet containing an explantory cover letter, the assessment battery, and a stamped, addressed return envelope. Follow-up was made either by postcard or telephone.

RESULTS

Analysis

Factor analysis was performed on the results of the needs assessment scale to better formulate the constructs of needs among military wives and to more fully explain the relationship between needs and psychological distress. The nine factors that emerged included: Role Expectations; Non-Spousal Relationships of Military Wives; Familial Relationships of Military Wives; Work Responsibilities/Opportunities of Husband; Independence; Separation Through War, Death, or Temporary Duty; Adequacy of Military Pay/Housing/Military Services; Leisure Opportunities Within the Military Community; Communication Between Military Establishment and Military Wife.

Two significant results were evidenced. Wives of Air Force personnel reported significantly more psychological distress on the BSI than did the female non-psychiatric norm group (t = 5.94, p .001). The second result was that wives of enlisted personnel reported significantly more psychological distress on the BSI than did wives of officers (t = 3.33, p. .01).

Supplemental Analysis

Correlationsl analyses were undertaken to examine the relationship between (1) subjects' years of education and report of psychological distress, (2) subjects' years of education and endorsement of needs, and (3) psychological distress and needs. Statistically significant correlational relationships were found between education and psychological distress and needs. Wives with higher levels of education reported the lowest levels of psychological distress, whereas wives with lower levels of education reported more psychological distress. This was also true for needs, that is respondents having more years of education reported fewer needs. A statistically significant, positive relationship was found between needs and psychological distress (p .001, r = .33).

An analysis of variance procedure was used to examine the relationship between number of friends at present location and (1) psychological distress and (2) needs. Air Force wives who reported higher numbers of close friends, at the present location, reported less psychological distress. Respondents with more than five close friends do not seem as dissatisfied with their lifestyle as do those persons reporting one to two close friends.

CONCLUSIONS AND RECOMMENDATIONS

A major finding of this study is the disproportionately high level of psychological distress among the wives sampled as compared to the normative group. Problems among military wives can be a deterrent to the performance and productivity of the husband. Dissatisfied wives have been found to be a strong factor in the husband's decision to leave the service and thereby contribute to the current retention problems faced by the Air Force (O'Keefe, 1984). Hence, the psychological distress and needs of military wives have far-reaching implications for the performance of the military mission.

Disproportionately high rates of psychological distress in conjunction with the correlation between psychological distress and needs suggest where some of the problems lie. Fifty percent or more of the present sample expressed

the following needs: (a) improved military benefits; (b) coping with separations from husband and from close relatives; (c) coping with the possibility of wars and husband's death; (d) desire for greater autonomy, independence, and self-sufficiency; (e) dealing with a boring environment; (f) coping with military role expectations -- integrating conflicting demands on personal time, energy, and long-term goals. Needs further delineated by comments respondents added to their surveys, painted a picture of the pilot training base as an environment in which decisions affecting wives are often made by others. According to these comments, prestige and consideration are accorded typically only on the basis of rank; personnel in military stores are often rude and inconsiderate; authority is sometimes used unfairly; wives do not have input or control over their lives. Several wives expressed the pain of moving and leaving behind a home, friends, and jobs they had grown to love. These characteristics suggest that it is difficult for wives to develop autonomy and independence within the military environment. That enlisted wives reported more psychological distress than officers' wives suggests that enlisted wives are more affected by the military environment and need different interventions than wives of officers.

Education was inversely related to the expression of both needs and psychological distress, suggesting that education may be a form of buffer against stresses of the military lifestyle. However, if the wife does not have a higher level of education before becoming a military wife, there often is not time or money to pursue a degree. Often, moves will interrupt the educational process, as with careers or jobs, making the effort difficult, or at best frustrating.

Support systems are extremely difficult in adjusting to present or new locations. The more friendships the military wife engages, the less she will be dissatisfied or maladjusted.

Recommendations that arise from these findings and conclusions begin with further research on specific problems at bases and strategies for coping with stress situations. The development of a mental health education program that addresses identification of problems and works on turning distress into a healthier approach would be welcomed. Military wives need programs in the area of intellectual, social, and vocational competencies. A series of topical workshops could be offered (through base chapel, clinic, base education office) on a variety of relevant topics (e.g., financial planning; dual career issues; coping with family separations; "survival skills").

In particular, it is recommended that a program be organized in which interested wives of the military community would be trained as peer leaders or peer counselors to work directly with other military wives under the supervision of the base psychologist. This would provide some built-in jobs for wives, leadership training, and experience, and would address an important community problem. Further, in the absence of a close friend, a peer counselor may serve some of the buffering purposes which friendships serve and may help an otherwise isolated military wife to develop more adaptive coping skills.

All of these recommendations could fit into a more comprehensive military community mental health delivery system. Military wives would then have more avenues to have needs met within the military system.

REFERENCES

Belt, J. & Sweeney, A. The Air Force wife: Her knowledge of and attitudes toward the Air Force. Paper presented at Military Testing Association Conference on Human Resources, San Antonio, Texas, October, 1973.

Bowen, G.L., Woolley, S.F., & McGaiphey, T.A. Department of the Navy Family Advocacy Program: Service need and service response. Phase I Report. Arlington: Department of Navy, February 1983.

Derogatis, L.R. Brief Symptom Inventory. Copyright, 1975.

Dohrenwend, B.P. & Dohrenwend, B.S. The problem of validity in field studies of psychological disorder. Journal of Abnormal Psychology, 1965, 70 (1), 52-69.

Jacobson, J.M. Effects of a structured learning experience on anxiety levels of women in midlife who are Air Force officers' wives. Unpublished thesis, San Jose, California: San Jose University, 1980.

MacIntosh, H. Separation problems in military wives. American Journal of Psychiatry, 1968, 125, 260-265.

Moskos, C., Jr. The American enlisted man: The rank and file in today's military. New York: Russell Sage Foundation, 1970.

O'Keefe, R.A., Eyre, M.L. & Smith, D.L. Military family service centers. In F.W. Kaslow & R.I. Ridenour (Eds.) The military family. New York: The Guilford Press, 1984.

Selyes, H. Stress without distress. Philadelphia: J.P. Lippincott, 1974.

Sobie, J.H. Wives clubs in the 1980s: A new perspective. Ladycom, September, 1981, 96-98.

Changing Marital Roles and the Support of Air Force Wives[1]

Philip Lewis, Dorothy Winther, and Angela Adan
Auburn University

Mickey R. Dansby
Leadership and Management Development Center
Maxwell Air Force Base

Abstract

Reflecting societal changes, Air Force wives are increasingly likely to be employed full time, view their own career as important, and marriage as a partnership between equals. These changes have raised fears that these "new women" will be less involved in and less supportive of Air Force life. Data examined in the present study from 811 enlisted Air Force members and their wives revealed that, contrary to expectations, employed wives are more supportive of Air Force life than unemployed wives. Similarly, among unemployed wives, those who are more positively disposed toward having their own careers are also most supportive of the Air Force. In contrast, wives with a "modern" or egalitarian view of marital roles have a less positive view of the Air Force than their "traditional" counterparts.

Introduction

The classical view of military life is that a person does not "take a job" in the military. Rather, he or she "joins" the military, becoming a part of a community and lifestyle that assumes a central role in one's life and, if married, the life of one's family (Carr, Orthner, & Brown, 1980). Although this military lifestyle can have its own special rewards, it can also make special demands, particularly on those with famlies. Due to temporary duty assignments away from home (TDYs), frequent moves, combat readiness, etc., the spouses of military personnel are often expected to adapt their own personal and vocational interests to the demands of military life. Spouses (typically wives) may have to assume full responsibility for household chores and child care during extended TDYs, and relinquish jobs to move with their spouses to remote or foreign duty locations where their own job opportunities may be limited. Because these demands are expected to be less for a non-working wife who sees her marital role as that of a homemaker supporting her husband's military career, there has been some concern about the impact of the changing roles of women in American society on the satisfaction and commitment of male military personnel. American women are increasingly likely to be employed full time, view their own

[1]The research described in this report was conducted under the sponsorship of the Air Force Systems Command, Air Force Office of Scientific Research.

careers as important, and marriage as a partnership between equals (Bernard, 1981). These changes have raised fears that these "new women" will be less involved in and less supportive of military life, with a resulting negative impact on the career commitment of their military husbands.

Are these fears warranted? Do the wives of military personnel who work and/or who have a more "modern" view of marital roles in fact have a less favorable view of the military? And are they less involved in the military lifestyle? An existing data set of the Air Force's Leadership and Management Development Center (LMDC) provided an opportunity to empirically address these questions. The data set consists of the matched responses of 811 Air Force enlisted men and their wives, where the male Air Force members had completed a job attitude survey, the Organizational Assessment Package (Short, 1985) and their wives the Air Force Family Survey (Flannery & Dansby, 1985). Previous research on the Family Survey (AFFS) had identified several factors relevant to the questions at hand (Lewis, 1985; Dansby, Lewis, & Winther, 1986). Two factors served as dependent variables in the current research, the wife's attitude toward the Air Force and her level of involvement in the Air Force lifestyle. The former was of particular interest, because previous research (Lewis, 1985) had found that a spouse's attitude toward the Air Force is a strong predictor of the Air Force member's job satisfaction and career intent. Independent variables were the wife's employment status, her reasons for working or not working, her attitude toward working ("career orientation") and her marital role orientation ("traditional" versus "modern").

Although a number of hypotheses are possible, the most parsimonious seemed to be that a wife's employment status, career orientation, and marital role orientation would not affect her attitude toward the Air Force but that they would affect her actual level of involvement in Air Force life. The rationale for these predictions is that while positive attitudes can be developed with little expenditure of effort, involvement does take effort. Since working wives typically have less discretionary time than do non-working wives (Brown, 1985), they are not expected to have the time to become very involved in the Air Force lifestyle. Similarly, wives who are interested in their own careers (rather than viewing a job as a necessary evil or interesting diversion) are expected to be less involved in the Air Force lifestyle by dint of their interest and involvement in their own careers. Finally, "modern" women are expected to be less invested in the role of wife as a support to her husband's career and more interested in their own independent development and lifestyle. This orientation is expected to be reflected in lower involvement in the Air Force lifestyle.

Method

Subjects

The 811 couples, Air Force enlisted men and their wives, were drawn from a census sample at three different air bases in the continental United States. Of the 811 wives, 331 were employed in civilian jobs, and 480 were not employed outside the home.

Procedure

Each hypothesis was tested separately by means of independent analyses of variance. All analyses examined two dependent measures, a factor score labeled "attitude toward the Air Force" and a second factor labeled "involvement in the Air Force lifestyle." The former represented the sum of six items (2, 5, 7, 18, 26, & 32) from the AFFS, and the latter consisted of the sum of seven AFFS items (1, 3, 4, 6, 9, 11, & 47), with the last item being reverse scored. Independent measures consisted of the working status of the wives (employed versus unemployed), reasons for working or not working, and two factor scores from the AFFS (career orientation and marital role orientation). Career orientation consisted of the sum of three AFFS items (24, 43, 44), and marital role orientation was the sum of two AFFS items (45 & 46). Career orientation was dichotomized on the basis of a median split with a score of 14 and above representing a positive orientation toward having one's own career and a score of 13 and below representing a negative career orientation. On a similar basis, a score of 9 and above represented a more traditional orientation toward marriage, and a score of 8 and below represented a more modern orientation.

Results

Two independent one-way ANOVAs were performed. The first analysis revealed a significant difference in attitude between the employed and unemployed wives, $F(1,809)=3.76$, $p<.05$. Contrary to our predictions, the employed wives indicated a more positive attitude toward the Air Force (M=29.08) than did the unemployed wives (M=27.89). No significant differences appeared between the two groups in terms of involvement.

Subsequent analyses revealed that the attitude of the employed wives toward the Air Force could not be predicted from their reasons for working. However, reasons for working proved to be a significant predictor of involvement, $F(6,307)=2.28$, $p<.04$. Those indicating that they worked to make use of free time showed a higher degree of involvement (M=39.57), while those working for personal or professional fulfilment showed a lower level of involvement in the Air Force lifestyle (M=33.46). Reasons for not working given by the unemployed wives did not prove to be significant predictors of either attitude or involvement.

Using the total sample of 811 military wives, separate two-way ANOVAs were conducted to examine the effects of career orientation and marital orientation on attitude toward the Air Force and involvement in the Air Force lifestyle. Results for attitude toward the Air Force demonstrated a significant main effect for both career orientation, $F(1,807)=12.35$, $p<.001$, and marital role orientation, $F(1,807)=6.32$, $p<.01$. Wives with a more positive career orientation indicated a more positive attitude toward the Air Force (M=29.14) than did wives less oriented toward a career (M=26.69). Wives with a more traditional the Air Force (M=28.80) than did those with a more modern perspective (M=27.04).

Similar results were obtained for involvement. Both career orientation, $F(1,807)=6.53$ $p<.01$, and marital role orientation, $F(1,807)=4.13$ $p<.04$, were significantly related to a spouse's involvement in the Air Force lifestyle. As predicted, wives with a more "traditional" view of marriage indicated more involvement (M=34.90) than did those with a more modern view (M=33.59). However, contrary to predictions, wives with a higher degree of career orientation indicated more involvement (M=35.06) than did those with less desire to pursue a career (M=33.43). There was no significant interaction between career orientation and marital role orientation in either analysis.

A separate examination of the employed wives' responses demonstrated no significant main effects or interactions for career orientation and marital role orientation scores when used to predict either attitude or involvement. In contrast, career orientation, $F(1,476)=8.76$, $p<.003$, and marital role orientation, $F(1,476)=4.53$, $p<.03$, were found to significantly predict attitude toward the Air Force in the subsample of unemployed wives. Those unemployed wives with a greater desire to pursue a career had a more positive attitude toward the Air Force (M=28.90) than did those with less of a desire to pursue a career (M=26.48). Unemployed wives with a "traditional " perspective of marriage showed a more positive attitude (M=28.56) than did those with a more modern perspective (M=26.82). In terms of involvement, only the career orientation of the unemployed wives served as a significant predictor, $F(1,476)=4.70$, $p<.03$, with those having a desire to establish a career showing more involvement (M=34.78) than wives who indicated less of an interest in pursuing a career (M=33.17). Career orientation and marital role orientation showed no significant interaction in terms of either the employed or unemployed subsample.

Discussion

Overall, the findings of the present study should be reassuring to those who fear that the entry of an increasing number of American women into the world of work could lead to an erosion of support for their military husbands. For wives of the enlisted Air Force members in our sample, those who work outside the home tend to have more positive attitudes toward the Air Force than their non-working counterparts. Similarly, wives who had a positive career orientation (think it important for a wife to have her own career), particularly those not currently employed, have more positive attitudes toward the Air Force than those who don't value working. And interestingly, although the opposite was predicted, those with a positive career orientation were more involved in the Air Force lifestyle. Overall, it appears that working wives and those non-working wives who have a positive career orientation are most supportive of Air Force life. Since the proportion of Air Force wives who work has increased dramatically over the past two decades, this should come as good news to the Air Force.

The bad news for the Air Force concerns changing marital roles. Wives who hold a more traditional view of marital roles have a more positive attitude toward the Air Force and, as predicted, they feel more involved in the Air Force lifestyle than those who have a more egalitarian view of marital roles. This effect appears to be most

pronounced for non-working wives, where non-working "modern" wives hold the least positive attitudes toward the Air Force. This could be increasingly important, if the recent trend toward more egalitarian views of marital roles continues.

Implications

Given that spousal attitudes toward the Air Force are a predictor of Air Force members' career intent, the findings reported herein may have implications for the retention of Air Force personnel. Because the employed wives of Air Force enlisted personnel apparently have more positive views of the Air Force, the trend toward greater participation of wives in the workforce should cause the military little concern. However, the paralell trend toward a more "modern" or egalitarian view of marital roles by American women may be more problematic. In the current study, it was the more traditional wives who had the most positive attitudes toward the Air Force, particularly in our subsample of unemployed wives. Assuming that the Air Force will continue to enlist men whose wives hold egalitarian views of marital roles, it will be important to try to find out the sources of these women's relatively lower regard for the Air Force. Are the sources of their less positive attitudes factors that can be addressed through family-oriented programs, or are they reflective of deeply held values that are not as likely to change?

References

Brown, C. (1985). An institutional model of wives' work decisions. Industrial Relations, 24, 182-204.

Bernard, J. (1981). Facing the future. Society, 18(2), 53-59.

Carr, R., Orthner, D.K., & Brown, R.J.,III. (1980). Living and family patterns in the Air Force. Air University Review, 31, 75-86.

Dansby, M.R., Lewis, P., & Winther, D. (1986). Family and work in the Air Force II:The Air Force Family Survey. Paper submitted to the Tenth Psychology in the Department of Defense Symposium, Colorado Springs, CO: United States Air Force Academy.

Flannery, P.A., & Dansby, M.R. (1985). USAF Family Survey:A revision of the USAF Spouse Survey (LMDC-TR-85-3). Maxwell AFB,AL: Leadership and Management Development Center.

Lewis, P. (1985). Family factors and the career intent of Air Force enlisted personnel (LMDC-TR-85-9). Maxwell AFB, AL: Leadership and Management Development Center.

Short, L.O. (1985). The United States Air Force organizational assessment package (LMDC-TR-85-2). Maxwell AFB, AL: Leadership and Management Development Center.

Adjustment to Overseas Living: The Case of the Insulated Sojourner

Deborah R. Richardson
and
Karen U. Millar
University of Georgia

Abstract

The present study was designed to determine correlates of adjustment of personnel assigned to overseas duty. Selection of potentially relevant factors was based on results of previous research on sojourner adjustment and consideration of the "remote" environment of the sample of respondents. Potential correlates included measures of interpersonal contact, loneliness, and job-related attitudes. The measures of adjustment were "global" life satisfaction and stress. Forty-two military personnel and sixty-four faculty completed a variety of questionnaires designed to measure these variables. Job-related attitudes were correlated with both measures for the faculty sample and with stress for the military sample. Interpersonal contact related to amount of stress reported by both groups of respondents.

Introduction

Previous studies have examined the psychological adjustment of people who live in a foreign country for a limited period of time (i.e., sojourners). Those studies (cf. Church, 1982) have focused primarily on foreign students, missionaries, and Peace Corps volunteers -- groups who "go native," live entirely within the foreign culture. In these cases, the most significant determinants of adjustment often are related to the ability of the sojourner to adapt to the foreign culture (i.e., learning the language, learning and accepting different norms for behavior). American business professionals and military personnel on overseas assignments, however, may be somewhat insulated from the foreign culture. They are likely to be living in a relatively structured environment (i.e., on a military installation) and to have frequent contact with other Americans. Thus, the correlates of their adjustment to overseas living may differ from those of the previously-studied sojourners.

The present study was designed to examine the adjustment of two groups of Americans living overseas -- college teachers and military personnel. Most of the members of these groups live and work among Americans, but this American environment is located in a foreign country. Thus, like other studies of sojourner adjustment, we focused on Americans living outside the U.S. Unlike previous studies, the present subjects were not

required to adapt to the norms and expectations of a foreign culture in that they could choose to spend most of their time among other Americans on the military bases. Thus our intent was to determine the correlates of adjustment for a particular type of sojourner -- the insulated sojourner.

Because most of the respondents in the present study were on a "remote tour" and were therefore unaccompanied by family or friends, loneliness and interpersonal contact were expected to be related to adjustment to living overseas. That is, those individuals who feel most isolated, who do not find close companionship, in the overseas environment, might experience more adjustment problems than those who feel less lonely or who develop close interpersonal contacts.

Church (1982), in his review of studies of sojourner adjustment, proposed that, "for individuals whose sojourn goals are significantly job related, adjustment may be strongly influenced by job conditions and satisfaction" (p. 550). Fredland and Little (1983) found that specific job-related attitudes were determinants of global job satisfaction among a large sample of military enlisted personnel. Thus, we included measures of specific job-related attitudes as potential correlates of adjustment.

We selected two indices of adjustment. A general measure was the respondent's indication of degree of satisfaction with life in recent months. Because physical and mental health problems are associated with adjustment difficulties, we also included a measure of stress.

This investigation is a preliminary step in an ongoing research project that examines adjustment among a larger and more diverse sample of sojourners.

Method

Respondents

The military sample consisted of 42 personnel stationed on two military installations in the Republic of Korea. Of these, 17 were affiliated with the Air Force and 24 were affiliated with the Army. Respondents were recruited through the first author's college-level introductory psychology classes. Data were collected in two ways: (1) questionnaires were completed by students in the classes (n=28); and (2) students who were unaware of the purpose of the study administered questionnaires to co-workers (n=14). Although the majority of the respondents were enlisted personnel (n=36), the data collected by the students included 3 officers and 3 civilians.

The college instructor sample consisted of faculty of a university that offers degree-related courses on military bases in Asia. Questionnaires were mailed to the 94 core faculty in the program at two different times. Thirty-seven completed

sets of questionnaires were returned from the first mail-out; 27 were returned from the second. Approximately half of the faculty change locations within Asia two or three times a year; the other half remains in one location. Most of the faculty live in quarters on military bases.

Questionnaires

Both military respondents and the first sample of faculty respondents completed the UCLA Loneliness Scale (Russell, Peplau, & Cutrona, 1980) and a Background Questionnaire designed by the investigators. The Background Questionnaire included items measuring demographic characteristics (e.g., age, sex, marital status), experience with and feelings about living outisde the United States, interpersonal contact, and loneliness. Military respondents also completed the Work on Present Job, Supervision on Present Job, and People on Present Job subscales of the Job Description Inventory (Smith, Kendall, & Hulin, 1975). Background Questionnaires completed by the first sample of faculty respondents included items dealing with their impression of their job. The second sample of faculty respondents completed only a few items from the Background Questionnaire and a measure of specific job-related attitudes.

Results

Descriptions and Comparison of Samples.

Of the military respondents, 29 were male, 11 were female, and 2 did not indicate their gender. Ages ranged from 19 to 50, with a mean of 26 years. The respondents had been in the military froml to 16 years (M = 5.15) and in Korea between 1 and 42 months (M = 8.19).

Of the faculty respondents, 73% (n = 46) were male, 27% (n = 17) were female, and one was unidentified. Ages ranged from 26 to 72, with a mean of 42.23 years. They had worked for their present employer between 1 and 23 years (M = 4.73). They had lived outside the United States an average of 4.94 of the last 10 years.

In order to examine differences between the samples, t-tests were calculated on those variables that were measured identically in the military sample and the first faculty sample. These analyses revealed that the military respondents reported greater loneliness (on both measures) and more stress-related problems than did the faculty respondents. Faculty had lived outside the U.S. longer than had the military personnel. There were no differences between the groups, however, in general job or life satisfaction. Means, standard deviations, and t-values for these comparisons appear in Table 1.

Table 1

Comparison of Military and Faculty Samples

	Military		Faculty		
	M	SD	M	SD	t
UCLA	41.33	7.327	35.378	10.473	2.439
Often lonely	2.900	1.150	2.324	1.082	2.408
Years Out U.S.	2.805	2.731	4.611	3.289	-2.632
Problems	6.405	3.914	4.297	2.885	2.223
Job Satis.	4.833	1.950	5.595	1.235	-1.686
Life Satis.	5.548	1.435	5.703	1.222	- .423

Correlates of Adjustment

Military Sample. All of the interpersonal variables and two of the four measures of job-related attitudes were significantly correlated with number of stress-related problems for the military sample. Neither of these sets of variables was consistently related to life satisfaction however.

Military personnel who reported relatively high numbers of stress-related symptoms also indicated that they were relatively lonely and they were not likely to report having a best friend. They were also relatively unhappy with their jobs, particularly with the work and the people on the job.

Faculty Sample. Although relatively lonely faculty in the second sample reported higher levels of life stress, no other correlations between interpersonal factors and the measures of adjustment were consistently significant.

Job-related factors seems to be more relevant to the adjustment of the faculty respondents than did interpersonal factors. Relatively stressed faculty reported that their job was stressful, was not fun, and that they were suffering from role overload. The highest and most significant relationships with job-related attitudes involved faculty reports of life satisfaction. Those faculty who reported relatively high life satisfaction also reported that their job was fun, challenging, not routine, and different. They reported that their jobs were relatively nonstressful and that they experienced relatively low levels of job overload. In addition, these faculty indicated high levels of "global" job satisfaction.

Discussion

These results provide support for Church's (1982) contention that sojourner adjustment may be related to job satisfaction and attitudes. Both groups of sojourners in the present study were living in Asia for professional reasons; thus their

adjustment is most clearly related to their experiences in the workplace.

The difference in significant correlates for the two samples supported our expectations. Interpersonal factors were relevant to adjustment for only the military sample. Correlates for the faculty sample, however, were concerned primarily with aspects of their jobs. Because the faculty actually sought out this job in an overseas environment, it is not surprising that these were the primary correlates of their adjustment.

The results reported here are based on a relatively small and unique sample of personnel on overseas assignments. Most of these individuals were on "remote tours." Moreover, the military sample may not be typical of military personnel assigned to overseas duty in other parts of the world: The Republic of Korea is still considered to be a third-world country (although that situation is changing rapidly); military exercises are relatively frequent and the possibility of renewal of hostilities is present; these bases are relatively small and isolated. Thus, one must be careful in generalizing the results of the present study. Nevertheless, the correlates we selected are not unique to the situations of the particular respondents in this study, so they may prove to be relevant to the adjustment of other "insulated sojourners" in other locations. We are continuing to collect data from military personnel and civilian employees in various locations in Asia and Europe in hopes of finding consistent correlates of adjustment to overseas living.

References

Church, A. T. (1982) Sojourner adjustment. Psychological Bulletin, 91, 540-572.

Russell, D., Peplau, L. A., & Cutrona, C. E. (1980) The revised UCLA loneliness scale: Concurrent and discriminant validity evidence. Journal of Personality and Social Psychology, 39, 472-480.

Smith, P.C., Kendall, L.M., & Hulin, C.L. (1975). The measurement of satisfaction in work and retirement. Bowling Green State University.

Fredland, J.E., & Little, R.D. (1983). Job satisfaction determinants: differences between servicemen and civilians. Journal of Political and Military Sociology, 11, 265-280.

De-mystifying Marriage for West Point Cadets
Effects of a Marriage and Family Course

Edna J. Hunter
John M. Wattendorf
United States Military Academy, West Point, New York

Abstract

An exploratory study of 42 West Point cadets who recently completed a one semester academic course in marriage and family was carried out. Responses indicated this course provided substantial new knowledge which was perceived by the cadets as very important to them, both personally and professionally. Results suggest a need for future military leaders to be educated in the area of marriage and family at the precommissioning level. Effective leadership can be quickly undermined if excessive time and energies must be focused on family issues. However, more definitive research is recommended.

Introduction

A stereotype of the recent past suggests that West Point tradition mandated a marriage ceremony at one of the Cadet Chapels immediately after the graduation ceremony, followed by a rice-strewn journey through arched sabers into marital bliss. So ingrained had been this tradition in years past, that a few cadets allegedly booked the Chapel for their weddings--and the Officers Club for the receptions-- even before they had chosen their brides-to-be. A longitudinal study of early career preparation, experiences, and commitment of female and male West Point graduates has been ongoing since 1980. In this study of 1980, 1981 and 1982 graduates of the United States Military Academy (USMA), factors which negatively affect Army family life have frequently been mentioned (e.g., lack of family time, separations, the spouse's career--usually the wife's, family mobility, and difficulties for dual military career marriages with the resulting dilemma over whether or not to have children). One interviewee cogently suggested, "I think before someone gets married,...(the military member) should set the person down and let her know exactly what she's getting into...you've got to make sure you show her the entire spectrum of what it would be like..." (Adams, 1985, p. 4-33).

To some extent, the USMA course in marriage and family accomplishes just that. It sets cadets down and informs them about a wide variety of marriage and family issues, with particular emphasis on families within the military context. There is a focus on the numerous responsibilities and potential problems in marriage which are often ignored when one is "blinded by the stars in one's eyes" when in love.

Love may be blind, but marriage is likely to have 20-20 vision!

Method

This exploratory pilot study examined the effects of having taken a course in marriage and family at USMA, on perceived changes in attitudes and future behaviors. The sample was a group of 42 West Point cadets who had completed a one semester course in marriage and family during their junior or senior years. Answers were sought to the following questions through cadets' anonymous responses on a paper and pencil questionnaire devised by the investigators:

. Do these cadets believe this course in marriage and family was important in relation to their future personal and professional lives?

. Do these cadets think this marriage and family course is important to the USMA curriculum?

. What new information did these cadets acquire from the marriage and family course which they had not previously known?

. What changes, if any, occurred in these cadets' attitudes/behaviors as a result of having taken the course (e.g., changes regarding timing of marriage, type of spouse, number of children, attitude toward careers for wives, and decision to make the Army a lifelong career)?

The research questionnaire asked specifically when the cadets planned to marry (e.g., within six months, within one year, or within five years), what new information they had acquired from the course, and what problem areas they could foresee in their own future marriages. Additionally, for each area examined, cadets were encouraged to comment freely on how the course in marriage and family related to that specific topic.

The Sample

Most cadets in the sample were seniors (91.9%); the remainder were juniors. Thirty-seven cadets in the sample (88.1%) were males and five (11.9%) were females. With respect to race, three-fourths of the group were white (76.2%), 14.2% were black, and 2.4% were Mexican-American. It should be noted these figures are not entirely representative of the total population of West Point cadets. The majority (52.4%) reported their religion to be Protestant; 25.7% were Catholic, and 2.4% belonged to the Church of the Latter Day Saints. Geographically, most cadets in the sample came from the Central (42.9%) or Eastern States (45.2%); few reported hometowns in other areas.

Results of the Study

Data collected were analyzed using simple frequency counts and percents. Although differences between male and females cadets were examined, no tests of significance were performed since the sample of females was too small to make meaningful comparisons (N = 5). Selected comments made by

cadets are reported to give a better understanding of these limited data.

Importance to Personal Lives

Cadets rated the importance of the marriage and family course for their future personal lives on a 5-point Likert-type scale, with "1" indicating "not at all important" to "5" indicating they believed the course had been "more important than any other course taken at the Academy." Mean score for the group was 3.7 on a 5-point scale, indicating, as a group, these cadets believed taking the course would contribute "a great deal" towards helping cadets lead a happier, more fulfilling personal life in future years.

Responses of the male cadets were similar to those given by female cadets on this item, although the latter viewed the course as slightly more important than male cadets did (Mean for females was 4.0; for males, 3.7). Because of the small number of female cadets in the group (N = 5), the difference might not hold for a larger sample. Ironically, one female cadet who rated the importance to personal life at the highest level of importance, added the comment that the course was "especially important for males." Still another cadet, a male, commented, "...the course made me feel more comfortable with the idea of marriage...I cannot imagine not taking this course....the thought scares me."

A variety of possible problems within their own future marriages was mentioned by the cadets, such as communication problems, power struggles, dual careers, and jealousies. One cadet commented, "...the bulk of our hangups and insecurities are related to sex role norms which are taught and may be changed." Another recognized "...it's tough to be married as a young Lieutenant in the Army." One of the female cadets added, "It opened my eyes to the great number of problem areas which need to be discussed...these little things can cause big problems."

Importance to Professional Life

Cadets also rated the marriage and family course as highly important for their future professional lives, although of slightly less importance than for their personal lives (3.5 versus 3.7, on a scale of 5). Males scored the course's importance to professional life slightly higher than did the female cadets (3.5 versus to 3.4, on a 5-point scale). With respect to the importance of the marriage and family course to the USMA curriculum, male and female cadets were in agreement -- both had mean scores of 3.6 on a 5-point scale -- indicating they believed the course should be required for most cadets.

None of the cadets indicated taking the course had made them want to leave the military service sooner. However, one female cadet indicated the course "...made me realize I might stay in (the Army) longer than I originally planned." Other

comments which illustrate the importance placed on taking the marriage and family course by the cadets included comments such as the following: "The course was very beneficial ... very worthwhile." "This course really opened my eyes to an experience which I took for granted would occur but had actually thought very little about. It made me think."

New Knowledge

Cadets derived a wide variety of knowledge from the marriage and family course which they had not previously known. The largest percents mentioned new information in regard to children and childbirth (31.0%), types of marital structures (26.8%), communication and negotiation skills (21.4%), and issues surrounding marital dissolution (19.0%).

Perceived Changes in Attitudes/Behaviors

Cadets were asked whether they believed their attitudes had changed or their future behaviors would change as a result of having taken the course in marriage and family. Specifically, we wanted to know whether they had changed with regard to the time at which they would marry, the type of spouse they would select, the number of children they planned to have, their attitudes toward wives working, or their decisions concerning staying in the Army.

By far the most frequently cited change in attitudes was concerning wives working outside the home. Almost half (47.6%) reported changes in attitudes. Since the wording of the question required a "yes" or "no" answer, it was impossible to determine precise direction of change in every instance. However, accompanying comments, for the most part, indicated a change toward the expectation of less traditional roles for wives. Cadets made such commments as, "It made me take a harder look at the importance of female aspirations;" "I'll be a little more encouraging for my wife to work;" "I don't mind my wife's working near so much as before;" "I would (now) choose a more independent woman who could support herself if I were unable to provide."

On the other side of the coin, some cadets' views concerning the necessity for having traditional roles for wives within military marriages appeared to be confirmed by the marriage and family course. This shift is evidenced in a number of comments: "I now know I need a wife who is willing to work, but will sacrifice (her career) to have children." "Yes, I've changed...I want a very traditional woman!" "I thought I could have a wife with a career and have one myself, plus children (but) that's not feasible."
One male cadet who reported having changed his attitude towards wives working, stated, "I do not want a wife with a career. She must be willing to support my career." This same cadet, in critiquing the marriage and family course as a whole, remarked, "...too much emphasis was placed on the negative aspects of an institution that has existed since the beginning of time. I would have preferred to hear about the

positive things that really make a marriage work."

Current Intentions Concerning Timing of Marriage

Cadets were questioned whether they planned to marry within six months, one year, or five years. The vast majority (95.2%) indicated they anticipated marriage sometime during the next five years. Most anticipated waiting more than a year (69.1%). Approximately 12 percent planned to take the big step immediately after graduation. However, over 14 percent reported having changed their minds concerning the timing of marriage after taking the marriage and family course. One cadet stated "...(I had) dated steadily before I took the course...just broke up with her -- 4 days after the course ended." Thus, for some cadets the course ended a relationship; for others, it appeared to clarify what cadets wanted in their marriages.

Discussion and Conclusions

The military has come a long way from the days when almost all soldiers lived in barracks, and a married junior enlisted soldier was the exception. Within all service branches, between 1952 and 1972 the percentage of enlisted men who were married showed a dramatic 77% increase. The increase for officers was smaller; 85% of male officers were married by 1973 (Segal et al., 1976). Today the majority (52%) of career soldiers, officer and enlisted, are married. Moreover, since 1960, there have been more Army family members than soldiers. Currently there are 384,000 spouses, 630,000 children, and 68,000 other dependent members, according to the latest Family Action Plan update (Department of the Army, 1985).

On a somber note, recent statistics suggest fully half the marriages begun in the United States in 1985 will end in divorce. Effective leadership is quickly undermined if leaders' time and energies are sapped by troubled marriages, either their own or those of their subordinates. Clearly there is a need for future leaders in all Armed Services to be educated at the precommissioning level concerning marriage and family. In conclusion, because this pilot study was limited in focus with a small unrepresentative sample, further research appears mandatory to clearly define the effects of instruction in marriage and the family on future marital histories and leadership experiences for officers from all sources of commissioning.

References

Adams, J. (1985). Early Career Preparation, Experiences, and Commitment of Female and Male West Point Graduate: Vol.III. West Point, NY: USMA.

Department of the Army (May, 1985). The Army Family Action Plan II. Washington, DC: Headquarters, DA.

Segal, D., Segal, M., Holz, R. Norbo, G., Seeberg, R & Wubbena. W. (1976). Trends in the structure of Army families, Journal of Political and Military Sociology, 4, 135-139.

AUTHOR INDEX

☆U.S. GOVERNMENT PRINTING OFFICE:1986—676-001/41033